

1993

The Influence of Teacher Behavior on the Distribution of Achievement in the Classroom: An Application of the Hierarchical Linear Model.

Leslie Sanford Arceneaux

Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation

Arceneaux, Leslie Sanford, "The Influence of Teacher Behavior on the Distribution of Achievement in the Classroom: An Application of the Hierarchical Linear Model." (1993). *LSU Historical Dissertations and Theses*. 5609.

https://digitalcommons.lsu.edu/gradschool_disstheses/5609

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.



University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9419863

**The influence of teacher behavior on the distribution
of achievement in the classroom: An application of the
Hierarchical Linear Model**

Arceneaux, Leslie Sanford, Ph.D.

The Louisiana State University and Agricultural and Mechanical Col., 1993

Copyright ©1993 by Arceneaux, Leslie Sanford. All rights reserved.

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

THE INFLUENCE OF TEACHER BEHAVIOR ON THE
DISTRIBUTION OF ACHIEVEMENT IN THE CLASSROOM:
AN APPLICATION OF THE HIERARCHICAL LINEAR MODEL

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Administrative and Foundational Services

by
Leslie Sanford Arceneaux
B.S., Louisiana State University, 1986
M.A., Louisiana State University, 1988
December, 1993

Copyright 1993
Leslie Sanford Arceneaux
All rights reserved

ACKNOWLEDGEMENTS

Alex Haley once said, "Anytime you see a turtle on top of a fencepost, you know it had some help." It is certain that this turtle would have never made the top of this fencepost without the support of many, many people. First, I'd like to thank Dr. Eugene Kennedy, my major professor, for taking me and this project on in mid-stream. His extensive knowledge of Hierarchical Linear Modeling and his willingness to share it with me made it possible to complete this study. Second, I was fortunate enough to have four very talented and dedicated researchers serve as the other members of my examining committee. Dr. Charles Teddlie deserves many thanks for his infinite patience and gentle (and sometimes, not so gentle) insistence that I should and could finish this enterprise. Dr. Abbas Tashakkori came to my rescue in what felt like my darkest hour when the HLM program wouldn't run. He has my gratitude for giving of his time so freely. The editorial advice of Dr. Diane Taylor was indispensable, and Dr. Terry Geske's ability to ask the pointed questions was also much appreciated.

I'd also like to thank my family for their unconditional belief in me and their unwavering confidence that I had what it took to finish what I had started. Thanks to my Dad for keeping me focused by forever asking me when I was going to finish "that damn dissertation." Thanks to my sister, Laurie, for just being available to listen

when I didn't think I wanted to finish. Thanks to my son, Garrett, (age 10 months) for only hitting the "Reset" button on the computer twice! Many thanks to his father and my husband, Shannon, for his help through balanced doses of encouragement and "reality checks." I also must thank the staff of the Broadmoor Baptist Preschool for being so understanding about the forced absences of the person discussed next.

More than anyone else, the credit for this project must be given to my mother, Sandra Sanford. From the inception of this study, she has stood ready to assist me in every way. Not only did she volunteer her time and energy to collect the data analyzed for the study, read every word of every version of this document (and there were many!), drive from Baton Rouge to Jackson, TN at a moment's notice to serve as chief cook and bottle washer for Garrett and Shannon, and provide dependable transportation for this manuscript to its final destination, but she also continued to be what she has always been - my best friend. It is to her that I dedicate this dissertation.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	vi
CHAPTER	
ONE	
INTRODUCTION	1
Background.	1
Statement of the Purpose.	5
Definition of Terms	6
Need and Significance of the Study.	7
Research Questions.	9
Scope and Limitations	11
Summary	13
TWO	
LITERATURE REVIEW.	14
Introduction.	14
School Effects and School Effectiveness	15
Teacher Effects and Teacher Effectiveness . .	25
Hierarchical Linear Models and Their Application to Issues of School and Teacher Effectiveness	32
Notes to Chapter Two.	60
THREE	
METHODS AND PROCEDURES	61
Overview	61
Population and Sample	61
Variables	66
Statistical Model and Analysis Strategy . . .	72
FOUR	
RESULTS	76
Overview	76
Descriptive Statistics	76
Results of the HLM Analyses	83
FIVE	
DISCUSSION AND CONCLUSION.	97
Discussion.	97
Limitations of the Study.	107
Conclusion.	111
REFERENCES	116
APPENDICES	
A	
SCORING KEY AND SHEET FOR STALLINGS CLASSROOM SNAPSHOT	127
B	
SCORING KEY AND SHEET FOR VIRGILIO TEACHER BEHAVIOR INVENTORY.	130
C	
STUDENT QUESTIONNAIRE.	134
VITA	136

ABSTRACT

The purpose of this study was to investigate the effects of classroom practices on the distribution of achievement within the classroom as well as on mean levels of achievement through the use of the Hierarchical Linear Model (Raudenbush & Bryk, 1986). The investigation focused on sixty classrooms - thirty from schools labeled as effective and thirty from schools labeled as ineffective. Data on teacher behaviors were gathered through classroom observations during which six dimensions of effective teaching were evaluated. These behaviors were interactive time-on-task, classroom management, strategies for monitoring student progress and providing opportunities to learn, strategies for presentation of content and questioning techniques, social/psychological environment of the classroom, and physical attributes of the room.

Once unconditional models were examined and their results indicated that there was significant variation in the class-level regressions, total battery scores from state achievement tests and the relationship between those scores and SES, measured by fathers' education, served as the dependent measures of two explanatory models. The first model dealt with the teacher behaviors in concert, while the second sought to isolate more specific teacher behaviors which might be associated with achievement and the relationship between SES and achievement in the classroom.

Results from the HLM analyses revealed a significant positive effect of effective teaching behavior on achievement. Specifically, classroom management was found to be highly significant in increasing class mean achievement. Interactive time-on-task and school type were found to have ameliorating effects on the within-class SES/achievement link, while increased effective teaching behavior, overall, and instructional strategies, specifically, seemed to be associated with a strengthened SES/achievement link within the classroom. It was suggested that this increase in association between SES and achievement implied an instructional emphasis by effective teachers which promoted excellence rather than equity.

CHAPTER ONE

INTRODUCTION

Background

In 1966, Coleman (see Coleman, et al., 1966) and his team of researchers conducted the Equality of Educational Opportunity Study (EEOS) which was intended to point to inequalities in educational opportunity based on race, social factors, gender or religion and relate these inequities to differences in student achievement. It was expected that such an investigation would highlight the impact of the school on student achievement. Surprisingly, the school level factors considered in the study did not show a significant relationship with achievement beyond the impact of family background and student ability. This now famous study challenged conventional wisdom that schools and their policies and practices have an ameliorating effect on the potentially negative impact of a child's socioeconomic background on his/her academic achievement.

In response to the Coleman, et al. (1966) study, researchers such as Weber (1971) and Edmonds (1978) began to investigate schools which, given their socioeconomic composition, should have exhibited low achievement but whose students performed at a high level. Such research points to Edmonds' and others' belief that inherent in the definition of an "effective school" is the notion that student

achievement must be equitably distributed among all children in the school. In other words, high mean school achievement alone does not necessarily make an effective school. All subsets of students must achieve at a high level in order for the school to be classified as effective (Levine & Lezotte, 1990). Parallels to the logic of this argument can be asserted at the level of the classroom.

If, by definition, effective schools lessen the impact of socioeconomic factors on achievement at the school level, then effective teachers should also distribute achievement equitably at the classroom level. In fact, it would seem that the equitable distribution of achievement at the classroom level would be a prerequisite for the same at the school level.

Recent teacher effectiveness research has been primarily focused on "process-product" relationships in which some particular teacher behavior or group of behaviors (the process) is seen to influence student outcomes – most predominantly achievement (the product). Although equity and quality, that is, high achievement in the classroom is clearly implied by the process-product paradigm, no studies have been carried out which directly investigate the effect of teacher behaviors and characteristics on the relationship between student achievement and socioeconomic status within the classroom. However, there has been movement by school effectiveness researchers in the recent past to include

teacher/classroom behavioral variables in typical school effects designs (Stringfield, Teddlie, & Suarez, 1985; Teddlie, Kirby & Stringfield, 1989).

Generally, process-product researchers have attempted to make their student samples as homogeneous as possible so that the relationships between teacher behaviors and the outcome of interest would not be obscured by the presence of subjects who differ widely from each other, as is prudent when correlating any two variables (Borg & Gall, 1989). This has been accomplished by analyzing levels of SES separately (Brophy & Evertson, 1976; Good, Ebmeier & Beckerman, 1978) or by including only one level of socioeconomic status in the study. As a result, process-product research has not been able to inform teachers about the ways that their behavior impacts relationships between variables (e.g., SES/achievement) within their classrooms.

Further, most of these studies have generated correlations between frequencies of teacher behaviors and class mean achievement or mean affective outcomes. Numerically this poses no problem, but conceptually, investigation of relationships to mean levels of outcome do not illuminate fully the effects of teacher behavior on student achievement (Burstein, 1982), particularly since, ultimately, outcomes of individual pupils of all backgrounds are of concern (Averch, Carroll, Donaldson, Kiesling & Pincus 1972). In other words, as has been intimated,

modeling the relationship of teacher behaviors to the distribution of achievement within the classroom may be of more substantive interest, particularly from an equity/quality perspective than correlations based on cross-class data.

For these reasons, Brophy and Good (1986) caution that it is important to reconsider teacher behaviors as parts of larger patterns occurring in particular contexts. These authors go on to say that future research in teacher effects will need to record data in such a way that within classroom relationships can be studied as well as between classroom relationships.

The recommendations of Brophy and Good (1986) call for present researchers to view teacher effects data from a more appropriate "multilevel" perspective. This appeal has been shared by a number of researchers both past and present (Burstein, 1982, 1989; Cronbach, 1976; Kennedy, Stringfield & Teddlie, 1993; Murnane, 1975; Lee, 1986; Raudenbush & Bryk, 1986, 1989).

A multilevel perspective implies a point of view in the examination of educational data. As stated by Burstein (1982),

One begins with the obvious notion that the process of schooling takes place in a multilevel (more precisely, hierarchical) organization involving, in its most concise form, three levels: pupils, classroom/teachers, schools. Pupils receive instruction, either individually or in groups, from teachers in classrooms; these classrooms, and the

pupils and teachers within them, are located within the schools (p. 1).

Because of these organizational features, the effects of schooling on individual pupil performance can exist both between and within the levels of the educational system. Since the conceptual models of the effects of schooling are multilevel (Barr & Dreeben, 1983), the statistical models must also be multilevel. Unfortunately, until recently, such a match between the conceptual and statistical models has not been possible. However, advances in statistical modeling, particularly Hierarchical Linear Modeling (HLM) offer the statistical tools needed to formulate and test more realistic models of schooling (Raudenbush & Bryk, 1986).

Statement of the Purpose

The purpose of this study is, using the HLM methodology, to investigate the extent to which particular teacher characteristics and behaviors in the classroom affect the magnitude of the relationship between socioeconomic status and achievement in third and fifth grade classrooms in two Louisiana parishes and whether this effect is more evident in schools labeled as effective.

Definition of Terms

Multilevel Data

Although Burstein (1980) distinguishes between the terms "multilevel" and "hierarchical" by saying that the former refers to horizontal configurations while the latter arises from the experimental design literature and refers to the nesting of one experimental unit within another, this study makes no such distinction. For the purposes of this study, the terms will be used synonymously and will refer to the nesting of one experimental unit within another unit. Specifically, students are nested within classrooms (teachers) which are nested within schools.

Hierarchical Linear Models

Although there have been several approaches developed in the recent past for dealing with multilevel data, (e.g., Goldstein's (1986) Multilevel Mixed Linear Model; DeLeeuw & Kreft's (1986) Random Coefficient Model; and the General Multilevel Linear Model (Mason, Wong & Entwisle, 1984), the method employed for analysis of the present data will follow that presented by Raudenbush & Bryk (1986) - the Hierarchical Linear Model (HLM). This approach to multilevel analysis draws heavily from the work of Lindley & Smith (1972) in explicitly laying out a hierarchical structure in which parameters estimated at one level become the outcome variables at the next higher level and in the

use of Bayesian estimation to arrive at model parameter estimates.

Need and Significance of the Study

Need

Ideally, classroom teachers should encourage academic progress of every student regardless of race, ethnicity or family background. Unfortunately, the relationship between academic achievement and socioeconomic status in classrooms is persistent (Lee, 1986; Teddlie & Stringfield, 1993; White, 1982). St. John (1970) goes so far as to conclude that the effect of social class on achievement is so powerful "that the influence of other background and school factors can be detected only if socioeconomic status (SES) is first neutralized through matching or statistical control" (p. 255).

It would seem that in order to hold out a greater hope for realizing the ideal of an equitable distribution of achievement in our public schools it is necessary to investigate what effect, if any, teacher demographic characteristics and teacher behaviors identified as effective have on the link between SES and achievement. Such investigations are of particular interest if it is found that these teacher behaviors can be manipulated. It is to this end that the present study addresses itself.

Significance

Studies linking socioeconomic status and academic achievement are abundant. White's (1982) meta-analysis of over 200 studies supports such a statement. Further, Brookover, Beady, Flood, Schweitzer, and Wisenbaker (1979) and Teddlie, Falkowski, Stringfield, Desselle, and Garvue (1984) have shown that 40-60% of the variance in mean school achievement can be accounted for by mean school SES, although there are multicollinearity problems inherent in these studies. Such studies span many grade levels, types of academic measures and types of socioeconomic indicators. Therefore, "this relation is so widely accepted that it is often cited as a self-evident fact" (White, 1982, p. 461).

Process-product research linking teacher behavior and student achievement is also a dominant theme in educational research. Brophy and Good's (1986) review of numerous studies in this field point to the fact that teacher behavior can influence academic achievement and that many of these behaviors can be manipulated to increase achievement. Process-product research has also provided insight into the dynamics of the classroom experiences of students from different socioeconomic backgrounds.

As can be seen, the inclusion of the central variables - teacher behavior, socioeconomic status, and achievement - in the present study is certainly not novel. However, to consider the relationships among these variables

from a hierarchical perspective is. To be sure, bivariate correlation of these variables and between-class comparisons in past research have done much to forward knowledge about how teachers behave and how children of various backgrounds achieve, but classroom research has seemingly ignored within class relationships, such as the SES-achievement link, which may vary systematically across classes due to the influence of differential teacher behavior. The investigation of these systematically varying within-class relationships is at the heart of the present study. Additionally, the study offers a significant departure from most of the current HLM research on teacher effects in that it seeks to study these relationships on the classroom level rather than the school level. Further, this investigation uses true behavioral data rather than archival data which has been used in most other HLM studies. It is hoped that with the use of such data, a more complete picture of the effects of teacher behavior on the strength of the SES-achievement link within classrooms may be drawn.

Research Questions

Specifically, the present study using the HLM methodology hopes to shed light on the following issues:

1. Is there a relationship between SES and achievement, on average, within classrooms? This will be accomplished by testing the hypothesis that the mean SES-achievement

slope (pooled within all classrooms) is zero. It is expected that this null hypothesis will be rejected.

2. Do mean achievement levels and SES-achievement relationships vary across classrooms? This will be accomplished by answering two questions: a) Does the variability in mean achievement across classrooms represent more than sampling error? and b) Does classroom variation in the strength of the SES-achievement relationship represent more than sampling error? Both of these questions involve the testing of variances. Question (a) can be answered by testing the null hypothesis that the observed differences among classrooms in mean achievement (the variance of mean achievement) could have occurred by chance alone. Question (b) can be answered by testing the null hypothesis that the variance among classrooms in the strength of the SES-achievement relationship could have occurred by chance alone. It is expected that both of these hypotheses will be rejected.

Once the above issues have been addressed using what can essentially be termed a "null" between-class model, the central, and more interesting, questions of the study may be pursued. These questions involve building and testing between-class models which seek to identify variability among classrooms with regard to mean achievement and strength of the SES-achievement relationship as a function

of school effectiveness and teacher characteristics and behaviors.

Specifically:

3. Is there a difference in the magnitude of the SES-achievement relationships between schools identified as "effective" and "ineffective"?
4. Does increasing time-on-task decrease the variability of mean achievement and decrease the strength of the SES-achievement link within classrooms?
5. Does the increasing quality of teacher behavior in the areas of instruction, climate and classroom management increase mean achievement and decrease the strength of the SES-achievement link within classrooms?

Each of these questions can be investigated by including the aforementioned variables in the between-class regression model which utilizes the mean achievement (the intercept) for each class and the SES-achievement link (a beta-weight) as dependent variables and then testing the null hypotheses that the regression coefficients for this between-class model are zero.

Scope and Limitations

As an outgrowth of behavioral psychology, much of the teacher effects research has described what happens in classrooms but seldom gives an explanation as to why. The present study shares this particular limitation in that it

does not attempt to establish causal relations in the experimental sense, but focuses instead on analyzing correlations. Nevertheless, the findings of this study may lead to the formulation of hypotheses that can be tested in the more restrictive experimental context.

Another limitation of the study is related to the data used for the analysis. Data were collected in only two elementary grades. Several researchers have suggested that elementary data may not be generalizable to middle or high school (Purkey & Smith, 1983; Virgilio, Teddlie, & Oescher; 1991). However, it can be argued that primary school is worth investigating because of the long term influence it appears to have on a child's academic career. Primary school is also the educational level at which almost all school effectiveness studies have been conducted (Brookover, et al., 1979; Mortimore, Sammons, Stoll, Lewis & Ecob, 1988; Teddlie, et al., 1984). Further, the influence of socioeconomic status in the primary grades is likely to be more distinct and more easily isolated (Kennedy, 1990).

The present study may also be limited in that the socioeconomic data were reported by the students. It has been suggested that young schoolchildren cannot accurately report paternal occupation (Mason, Hauser, Kerckhoff & Poss, 1976). However, each student's responses were verified through the classroom teacher.

Lastly, although the schools that were utilized in the study were chosen through a regression procedure which is explained in later chapters, it was still necessary to obtain permission from each school's principal to access the teachers and students within those schools. In this sense, the school's participation in the study could be considered voluntary. Although comparisons were made to assess the representativeness of the sample, caution may be advised in generalizing the findings to other non-volunteering schools.

Summary

As has been stated, the relationship between a child's social class and his achievement is one of the more consistent (and persistent) findings in the educational research. Although the present study does not attempt to identify causal agents linking family background to achievement, the effort is made to identify class (teacher) level factors which may impact this relationship.

The following chapter will present a review of the relevant literature and describe in detail the nature of the model which will be studied. The third chapter will outline the sampling, instruments, and statistical methods employed in the investigation. The final two chapters will present results, conclusions and implications of the study.

CHAPTER TWO

LITERATURE REVIEW

Introduction

Teachers and the researchers who study them have long been interested in increasing the academic achievement of students. As a result, much research over the past thirty years has been directed at trying to identify particular teacher attributes and behaviors which have some bearing on bolstering low achievement or maintaining high achievement, usually defined as scores on a standardized instrument.

In its earliest phase, the research called, both directly and indirectly, for the equitable distribution of high achievement within the classroom regardless of ability, ethnicity or socioeconomic status. As this work continued to evolve, the inclusion of "context" variables in research designs brought about a more expanded view of effective schools and teachers which posed questions involving efficiency, as well as equity. In short, the ideal classroom, regardless of context, is one in which "input variables," such as, student background, attitudes, or personality, would be insignificant in their prediction of successful achievement. Such a classroom would make it possible for schools to become what they have been historically portrayed - the "great equalizers of the conditions of men" (Greer, 1972).

Investigation of school and teacher effects on student achievement has been conducted under various rubrics. It is the intention of the following review to explore these studies with regard to their methodologies, as well as their findings. The review will then discuss the Hierarchical Linear Modeling (HLM) methodology to be used in the present investigation. It is believed that the multilevel perspective of HLM will be useful in revealing relationships among background, process and outcomes across and within the levels of the educational system not previously evident with other paradigms.

School Effects and School Effectiveness

The Educational Production Function

As defined by Geske and Teddlie (1990, p. 194), "an educational production function expresses mathematically the relationship between school inputs (e.g., socioeconomic factors, student characteristics and teaching personnel) and school outputs (e.g., gains in achievement results, growth in cognitive skills, and affective behavior)." The most prominent of the early production function studies is the Equality of Educational Opportunity Study (Coleman, et al., 1966), better known as the Coleman Report.

The U.S. Office of Education (Mosteller & Moynihan, 1972, p. 4-5) commissioned James Coleman and seven other researchers to undertake a study of public schools directed

at documenting differences in educational opportunity which were based on race, religion or national origin. The study involved some 570,000 students and more than 60,000 teachers and administrators and "became a central model for school effects research for the next twenty years" (Kennedy, 1990). The study showed clearly that the bulk of school characteristics considered had only a minimal effect on student achievement beyond the impact of family background. However, of the school-related variables examined, teacher characteristics had the greatest impact.

The controversial findings of the Coleman Report launched many challenges from other researchers on both conceptual and methodological grounds, bringing on a flurry of research activity. It could be said that the knowledge we now possess about schools, teachers and their effects on achievement may have been much delayed had Coleman's work produced the expected outcomes.

From a methodological perspective, Hanushek and Kain (1972) questioned the relationship of the sampling units to the analysis and inference levels of the study. Additionally, it has been suggested that because the questionnaires used to document school resources only addressed the presence or absence of resources and did not consider quality, their utility was limited (Armor, 1972).

These criticisms notwithstanding, many of the studies which attempted to rebut the findings of the EEOS (Hanushek,

1972; Katzman, 1971; & Levin, 1970), although improved in the areas of specificity of the production function and methodological and statistical techniques, were not completely successful (Geske & Teddlie, 1990). Hanushek's (1986) meta-analysis of 147 educational production function studies continued to point to the absence of significant effect for such teacher input variables as teacher/pupil ratio, teacher experience, teacher salary and teacher education.

Geske and Teddlie (1990) point out that the educational production function research paradigm has been consistently unable to show significant teacher effects because of its "inability to capture 'skill' differences across teacher inputs." They go on to conclude:

The concept of skill differences acknowledges that teachers with the same measured attributes (e.g., years of experience, college degrees, state teaching certificates) may exhibit quite different teaching styles and methods in the classroom, and that some of these differences in behavior or technique may be important determinants of school achievement. Although teachers may be the most important input variable in the school process, the measures of teacher attributes used in production function studies may not adequately detect or capture those teacher qualities or behaviors that systematically count. (pp.198-199)

Murnane and Nelson (1984) see such variation in teaching practice as unavoidable and much desired. They write:

...effective teaching requires information about the skills and personalities of students and about how students interact that can only be obtained during the classroom process. [Therefore] what one teacher does in applying a particular broadly defined method will

diverge, often considerably, from what another teacher does. (pp. 362-363)

These characterizations of teaching techniques as largely idiosyncratic activities serve only to highlight the notion that students organized as classroom groups nested, as it were within teacher, can and do receive quite variable instruction from other such aggregates even within the same school. Such a nested arrangement of students within educational "treatments" requires a quite different type of conceptualization from that forwarded by the educational production function to detect those school factors and teacher behaviors that do impact achievement.

Teacher Effects as Part of School Effectiveness Studies

As a result of criticisms like those stated above, other educational production researchers have included additional educational effects in their work in an attempt to show that school inputs do not predict student outcomes independent of school process. While Coleman, et al. used only school level, archival data in the 1966 study, Brookover, et al. (1979) added climate questionnaires that assessed learning environments in schools and classrooms (such as teacher assessments of schools' educational climates). He and his colleagues found that although climate variables were highly correlated with initial characteristics of students and teachers, these variables explain as much variation in achievement as do school input variables.

Some years later, Teddlie, et al., (1984, 1993), replicated Brookover's work. Five orthogonal second-order factors - two SES factors and three school climate factors - emerged from these researchers' use of a second order factor analysis. It was found that the three school climate factors, taken together, predicted more variance in student achievement (39 percent) than did the SES factors (34 percent), thus confirming and extending Brookover's results. A recent reanalysis of the Teddlie, et al., data by Kennedy, Stringfield and Teddlie (1993) found similar results using the Hierarchical Linear Modeling methodology.

Rutter, Maughan, Mortimore, Ouston, and Smith (1979) conducted a three year study in twelve secondary schools in England. The school processes of interest included academic emphasis, rewards and punishments, teacher actions in lessons, conditions of learning for students and student responsibility and participation in the school. Data on processes were derived from pupil response to questionnaires, interviews with teachers and classroom observation. In general, Rutter and his fellow researchers found that despite large differences in input characteristics, there were substantial and statistically significant differences between school outcomes and that these differences were systematically related to school characteristics (e.g., identifiable factors in academic emphasis, teacher behavior, etc.).

Other studies (Teddle, Kirby, & Stringfield, 1989; Teddle & Stringfield, 1993; Virgilio, Teddle, & Oescher, 1991) added teacher classroom behavioral variables more often associated with teacher effects research to typical school effects designs. Specifically, Teddle, Kirby & Stringfield (1989) reported that teachers in effective schools outscored teachers in ineffective schools on nine of ten effective teaching dimensions, including time on task, independent practice and high expectations. They further found that teachers in effective schools exhibited less variation in their behavior than their counterparts in ineffective schools.

Expanding the Teddle, et al. (1989) study, Virgilio, Teddle, and Oescher (1991) added a "typical school" level of effectiveness to be investigated and found that the three levels of effectiveness had "distinct effects" (p.162).

These researchers reported:

...Teachers in more effective schools consistently exhibited the following behaviors: a) began classes on time, b) used transition time effectively, c) used a positive approach to managing student behavior, d) focused students back "on task" when necessary, e) used above average instructional strategies in lesson presentations, f) displayed student work in the classroom and g) established a positive learning environment. (p. 162).

Further replicating the results of the Teddle, et al. (1989) study, Virgilio (1991) and her colleagues also found teachers in more effective schools behaving more similarly to one another than those in less effective schools.

Additionally, this study included the investigation of grade level differences (elementary vs. junior high) and noted differences in mean and deviation scores which suggest differential school processes at operation in those two levels.

Although these studies have included teacher level variables in their analyses and have produced many useful findings, their single level analyses have not allowed for the observational dependencies which are inescapable when dealing with intact classroom groups which are nested within schools. A notable exception, however, is the investigation conducted by Teddlie, Kirby and Stringfield (1989) which included analyses at both the school and classroom levels in an attempt to better model the data. The present study will account for these dependencies using the HLM methodology which allows for the simultaneous consideration of both within-group and between-group components of educational relationships.

Norms for School Effectiveness: Equity, Quality and Efficiency

Having traced the development of the school effectiveness research paradigm, it seems appropriate to discuss these studies from an expanded, political perspective. Wimpelberg, Teddlie, and Stringfield (1989) have summarized the "post-Coleman" (p. 82) effective schools research and have proposed that this research can best be

categorized as being of two distinct genre - one of equity and one of efficiency. They go further to suggest a third era which should serve both interests. It is within the boundaries of these authors' framework that issues of equity, quality and efficiency can most easily be discussed.

Ronald Edmonds, an advocate for the equitable distribution of achievement in American schools, defined "equity" as "a simple sense of fairness in the distribution of the primary goods and services that characterize our social order" (1979a, p. 2). He believed that one of the primary services of our social order was education which he defined as the "early acquisition of those basic school skills that assure pupils successful access to the next level of schooling" (1979a, p. 2). Edmonds was convinced that children of the poor were being systematically denied this service. Edmonds (1979a) stated:

Schools teach those they think they must and when they think they needn't, they don't. That fact has nothing to do with social science, except that the children of social scientists are among those whom schools feel compelled to teach effectively. There has never been a time in the life of the American public school when we have not known all we needed to in order to teach all those whom we chose to teach. (p. 3)

As a result of these views, Edmonds, among others, took the Coleman Report (1966) findings of the school's seeming nonrelevance to student achievement as a challenge and set out to prove otherwise. This was done by searching for schools for the "urban, poor" which did provide higher achievement in spite of socioeconomic and family factors.

Edmonds' search was an admirable pursuit - and a successful one. However, it is interesting to note that Edmonds and his followers used the education received by middle-class children as the standard by which to judge the achievement of children in lower socioeconomic levels. This view assumes that the education all middle-class children receive is of a high quality. Such an assumption is erroneous in that middle-class children have also been found to receive education that is of low quality (see, for example, Stringfield, Teddlie & Suarez, 1985). Therefore, merely equalizing educational quality may not be sufficient in assuring effective schooling. Nonetheless, this assumption brought about a series of investigations which exclusively studied schools serving children of low socioeconomic status (Brookover, et.al, 1979; Edmonds, 1979b; Klitgaard & Hall, 1974; Weber, 1971).

Wimpelberg, et al. (1989) point to the emergence of a second era of school effectiveness research as continued work in the area of school effectiveness discovered that the "formula" for effective schooling espoused by equity researchers was not replicated by research on secondary and higher SES schools (Hallinger & Murphy, 1985; Miller & Yelton, 1987). This era is one which Wimpelberg, et al. (1989) report shifted value categories from equity to "efficiency."

This value shift is described by Wimpelberg, et al. (1989) not as a substitute for equity but as a "by-product" of a research shift to concerns for effectiveness (i.e., quality) in all schools as well as concerns for controlling fiscal resources. However, the researchers' assertion that this shift might have been part of a broader based political movement toward "excellence" is far more interesting. William Boyd (1987) states that the rising level of education in the general population, particularly the middle and upper-middle classes has had the effect of making these citizens "...more sophisticated, discerning, and demanding consumers of educational services...This makes the public and especially the very highly educated upper-middle class, increasingly quality conscious and unwilling to accept mediocre schooling services" (p. 86)

As a result, Wimpelberg, et al. (1989) propose that the future of effective schools research take on the dual purposes of equity and efficiency. They conclude:

Effective schools research that is context-sensitive may be important to the improvement of schools for the poor and the preservation of public schools for the middle class. ...Research that is sensitive to multilevel effects, in particular the effects of individual practices and adult attitudes on children of varying SES backgrounds *within classrooms*, can preserve something of the equity impulse and may be a link to stabilizing the middle class population in already socio-economically integrated public schools. (p. 102)

This type of multilevel analysis of within classroom relationships is precisely the endeavor of the present study.

Teacher Effects and Teacher Effectiveness

The next research area of interest is the study of teacher effects which has developed separately from the study of school effects over the past 20 years. The process-product paradigm which attempts to honor the individualistic nature of teacher behaviors by investigating actual teacher behavior in the classroom is the focus for the following section. However, it should be noted that this paradigm largely ignores the hierarchical structure of educational data.

Correlational Process-Product Research of Teacher Effects

What do good teachers do? The answers to this question are complex, and in some ways elusive (Olson, 1988). Nevertheless, answers - if only partial ones - have been generated by the process-product research paradigm. In general, the study of "teacher effects" (as termed by Brophy & Good, 1986) has evolved through six paradigms of which "process-product research" is one of the more recent (Borich, 1986). In this paradigm, classroom/teacher characteristics are viewed as the processes which impact the product, student outcomes. This belief in a causal connection between teacher behavior and pupil behavior is reflected in the inclusion of behavioral interactions in the operational definitions of variables - a feature quite different from previous paradigms that related effective teaching to general personality characteristics of the

teacher which were assumed to impact the general disciplinary climate of the classroom. It is interesting to note that a similar evolution occurred in the research of school effects with the substitution of behavioral considerations for survey items, as noted in the previous section of this review.

Brophy and Good (1986) have summarized the findings of this research in this way. "Achievement is maximized when teachers emphasize academic instruction as a major part of their own role, expect their students to master the curriculum and allocate most of the available time to curriculum-related activities" (p. 360). Further, effective teachers present information actively and clearly, are task oriented, and move at a relatively fast pace (Good, 1983; Good, Grouws, & Ebermeier, 1983; Smith & Land, 1981). They limit student decision making choices, socialization, and nonacademic activities (Good, 1983; Medley, 1977; Stallings, Needels, & Stayrook, 1979).

Through a careful analysis of the literature on teaching behaviors, Teddlie, Virgilio, and Oescher (1990) identified three major skill areas of teacher effectiveness, all of which are reflected in the above descriptions: classroom management, instruction, and classroom climate.

Classroom Management Research. Good classroom management leads to more learning; poor classroom management leads to less learning (Coker, Medley & Soar, 1980; Good,

Grouws & Ebmeier, 1983; Good & Brophy, 1987). This point of view can be found in much of the professional literature (Anderson, Evertson & Brophy, 1979; Medley, 1977; Rosenshine, 1976; Soloman & Kendall, 1976; Stallings & Kaskowitz, 1974). As Brophy and Good (1986) state:

The largest adjusted achievement gains occurred in classes of teachers who were well organized, who maximized the time devoted to instruction and minimized time devoted to preparation, procedure, or discipline, and who spent most of their time actively instructing the students and monitoring their seatwork. (p. 350)

In addition, Soar and Soar (1979) found that effective teachers limited pupil freedom of choice, restricted physical movement, allowed fewer disruptions, controlled pupil behavior and talked more - but only up to a point of diminishing returns where too much teacher control had a negative effect.

Effective teachers allowed fewer disruptions of all types than less effective teachers according to Good, Grouws and Ebmeier (1983) and interrupted less what they were doing for matters related to student misconduct (Larrivee & Algina, 1983). Good teachers prevent student misconduct (Good, 1983) by anticipating problems thus limiting opportunities for students to be disruptive (Kounin, 1970). One way effective teachers limit disruptive opportunity is by using less time for transitions (e.g., going from one activity to another) and clearly communicating when such transitions were taking place (Brophy, 1979; Doyle, 1984).

Key to establishing such smooth transitions and successful management in general is establishing clear and consistent routines so students know what to do and when (Brophy, 1983; Rogoff & Wertsch, 1984). Emmer, Evertson and Anderson (1980) showed that the seemingly automatic, smooth functioning of the classrooms of successful managers resulted from the fact that not only were students told what was expected, but also had correct procedures modeled for them. Such routines make it possible to have a number of activities going on simultaneously (Doyle, 1985). On the other hand, rules should be kept to a minimum and should have a convincing rationale (Good & Brophy, 1987).

Although effective teachers exhibit more appropriate behaviors in the area of classroom management, it is obvious that effective management techniques should only pave the way for effective instruction.

Research on Effective Instruction. That students learn what they spend time on is hardly a startling research finding. Corno (1979) concluded, "Time becomes the most immediately promising focal point in the effort to improve achievement." Further, spending time on noneducational tasks leads to lower achievement (Larrivee & Algina, 1983). Rosenshine (1976) found that essential for achievement is time spent engaged in relevant content.

However, because time is not related to achievement in any simple or direct way (Karweit, 1983), merely allocating

time to a subject is not as important as what happens during that time. Brophy and Good (1987) have said, "Some teachers who allocate less time for a subject have considerably higher rates of academic learning time because they involve students more in appropriate tasks" (p. 36) and because they clearly explain things.

Brophy and Good (1986) in their comprehensive review of process-product research found clarity of presentation to be "one of the more consistent correlates of achievement" (p. 354). The structuring of optimal lessons with regard to clarity include the following: providing overviews of what is to be learned, outlining content, signaling transitions between parts, focusing on main ideas, relating new information to what has been learned previously, giving examples, summarizing subparts and reviewing main ideas at the end (Good & Brophy, 1987).

Although not directly related to clarity, another aspect of effective instruction is structuring lessons at an appropriate pace. A relatively rapid pace is best (Good, 1983; Good & Grouws, 1975), but not at such a rapid pace that the teacher does not have enough time to think and adjust instruction as needed. It has also been suggested that faster pacing is appropriate in dealing with lower level skills, but that some wait time between teacher questions and student answers is beneficial to achievement with regard to higher level objectives (Tobin & Capie,

1982). Further, it has been found that students who were involved in interactive teaching achieved at a higher rate than those who were engaged in seat work (Brophy, 1982; Stallings & Kaskowitz, 1974).

Classroom Climate. The overall climate of the classroom has been found to be another teacher-controlled aspect of instruction which affects achievement. Teachers who focus on authority and discipline have been found to be ineffective in promoting academic achievement (Brophy, 1983; Flanders, 1970). Yet, an excessively warm climate has not been related to increased achievement either (Rouk, 1979). The general conclusion has been that a neutral climate is best where teachers are business-like, enthusiastic, non-evaluative, objective, relaxed and believe that their students are capable of learning (Doyle, 1985; Brophy & Evertson, 1976; Brophy & Good, 1986; Good, 1983; Soar, 1968; Wright & Nuthall, 1970).

Experimental Process-Product Research on Teacher Effects

Thus far, the studies of teacher effectiveness considered have been correlational in nature. However, the process-product research paradigm has also produced experiments which came about as a result of the accumulation of some stable findings and in response to a call by Rosenshine and Furst (1973) for work on the "descriptive-correlational-experimental loop."

Gage and Needels, (1989) reviewed 13 experiments in the area of teacher effectiveness which met eight criteria:

1. Regular teachers were studied.
2. The regular curriculum was used.
3. A whole school term or year was devoted to the experiment.
4. Random assignment of teachers (or some instances, schools) was often used in forming the experimental and control groups.
5. The independent variables were derived (in large part, at least) from the findings of prior correlational studies of process-product relationships.
6. The teachers were observed.
7. Measures of implementation were obtained.
8. Measures of student achievement, attitude, and conduct were obtained.

All of the thirteen experiments included both treatment and control groups. Treatment groups were trained in a set of teaching practices, while control groups were not.

Overall, the teacher education programs brought about substantial increases in the use of the recommended teaching practices in all but one of the studies analyzed.

Additionally, these programs, based substantially on the results of previous correlational process-product studies, tended to improve mean class achievement by about 20 percentile ranks. According to Gage and Needels (1989):

In view of the brevity of the typical teacher education program used in these experiments, the results are substantial. More extended and thorough teacher education programs might produce even stronger results....The general conclusion inferable from these

experiments is not that a particular kind of teaching is generally better than another. Rather, it is that teaching practices identified as promising in a given grade level and subject matter through observational and correlational studies turn out to have some causal efficacy in that grade level and subject matter. (pp. 282 & 284)

There can certainly be little doubt that the process-product paradigm has yielded much in the way of knowledge about teachers and effective teaching. However, methodologically, this paradigm may have missed certain important relationships due to its reliance on single level analyses. Although the work done by school effectiveness researchers has attempted to give the multi-level structure of schooling more emphasis with the use of different units of analysis in the same study (Teddlie, Kirby, & Stringfield, 1989), teacher behaviors and their effects on the distribution of achievement within classrooms has still been inadequately explored.

Hierarchical Linear Models and Their Application to Issues of School and Teacher Effectiveness

In an address made at a conference on Data Aggregation Problems in Educational Research, Cronbach (1976) made the assertion that:

The majority of studies of educational effects... have collected and analyzed data in ways that conceal more than they reveal. The established methods have generated false conclusions in many studies. (p. 1)

This provocative assertion was grounded in a concern held by Cronbach and others for what typically had been single-level

analytical approaches for studying educational data. Single-level analyses, in many cases, failed to capture the complexity inherent in the study of educational effects. The ensuing search for "multilevel analysis" strategies has been founded on a "conception of multifaceted interconnections and effects between individuals and the social settings in which they are embedded" (Burstein, Kim & Delandshere, 1989).

More specifically, at the heart of all modern attempts to analyze multilevel educational data are two questions:

- 1) How can phenomena of interest be appropriately modeled, given that individuals (i.e., students) are found in naturally occurring social groups (classrooms, schools)? and
- 2) What analytical strategies will allow a disentangling of effects from multiple sources so that examination of the relationships among individuals and their groups and the implications of those relationships for understanding particular phenomena in the social setting of "school" or "classroom" is possible?

The purpose of this section of the literature review is to delineate methods for at least coming to terms with these questions. The first portion will define "multilevel analysis" and offer a general model for dealing with multilevel data. Succeeding portions will outline not only ways that such analyses have been and can be used in research on the effects of schools and teachers, but also

ways in which these analyses fall short of dealing with certain substantive and methodological concerns.

The Model

Before presenting the technical aspects of multilevel analyses, it is important to understand the concepts behind the model. The term "multilevel analysis" is used generically in the literature to refer to any set of analytical procedures that involve data gathered from individuals and from the social structure (in educational contexts, the classroom or school) in which the individuals are embedded, or nested. Accordingly, these data are analyzed in a way which models this multilevel structure. Because they better reflect the complexity of the phenomena at work and the inherent design structure in gathering the data, multilevel analytical models are desirable. In short, "they fit the data better" (Burstein, et al., 1989). Most of the time, it is not unreasonable to expect that the relationships between student characteristics and student outcomes will vary across classrooms. Further, this variability can be seen as systematic in part due to the impact of some set of group (macro) level attributes, such as, classroom organization, school discipline policies, principal leadership style, and teacher confidence. It can be said that a "cross-level interaction" (Burstein, 1989) exists such that within-school or within-classroom (micro-

level) relationships vary systematically across schools or classrooms.

Statistically speaking, it is understood that knowledge of relevant explanatory variables, both in terms of measurement and specification, and their impact on the micro-level relationships will be imperfect. Therefore it is anticipated that the group-level relations will have a stochastic element. Thus, there will be fixed effects (the effects of macro-level explanatory variables) and random effects (unmeasured or unexplained variability) associated with the macro-level contexts. Models including such random effects have been proposed under a variety of names: mixed models, variance component models and hierarchical linear models. Regardless of label, the intent is to identify the antecedents of student performance, or some other criterion, and estimate the magnitude of their effects.

Sampling. The typical HLM study involves two-stage sampling. A sample of schools (random, representative, or convenience) is drawn and students are sampled randomly (or on a stratified random basis) from within schools. Alternatively, a sample of classrooms is chosen (either randomly or exhaustively within a sample of schools) and all the students within the classrooms comprise the total study sample. In either situation, the resultant data will have dependencies among observations within the first-stage sampling unit. In short, there will be correlations among

individuals' scores in the same group. Therefore, it is necessary to recognize in the analysis of individual student results that students in the same school or classroom share common experiences which make their results more homogeneous than those of a random sample of students drawn from the population of all schools. The hierarchical linear model (HLM) makes provision for such dependencies.

Variables of Interest. Variables in most cross-sectional investigations are of three types:

- a. measures of student characteristics (attributes upon entry to the classroom or at some point prior to a period of instruction - e.g., student background, abilities, prior knowledge, attitudes, personality).
- b. measures of aspects of the teacher, classroom, school or program and the students' experiences in them (e.g., instructional resources, organization, content coverage and emphasis, atmosphere, teacher and school attributes.)
- c. measures of student characteristics (outcomes, performance) at a point following a period of instruction (e.g., test performance, marks, attitudes, motivation) (Burststein, et al., 1989, p. 238).

Again, these data are inherently multilevel because student-level attributes and the group-level characteristics in which they are nested are both investigated. This investigation is accomplished through the use of both a micro-level (within-group) and a macro-level (between-group) equation.

Basic Hierarchical Linear Model. Generally, the micro-level (within-group) equation for applying HLM to research

on school or teacher effects specifies the relationships among various student level characteristics, X_{ijk} , and some student outcome, Y_{ij} .¹

$$Y_{ij} = \beta_{j0} + \beta_{j1}X_{ij1} + \beta_{j2}X_{ij2} + \dots + \beta_{jk-1}X_{ijk-1} + R_{ij} \quad (1)$$

where

Y_{ij} is the outcome score for student i in context j ;

X_{ijk} are values on a set of student level characteristics for individual i in context j ;

R_{ij} represent random error in the student level equation;

and,

β_{ijk} are regression coefficients that characterize the structural relationships within context j ;

for,

$i = 1 \dots n_j$ students within context j ;

$j = 1 \dots J$ contexts; and

$k = 1 \dots K-1$ independent variables in the first stage model

The assumptions associated with equation (1) are as follows:

- i. For each context the values of X are fixed - estimation is conditional on this specific set of X 's.
- ii. The disturbances (random errors), R_{ij} are approximately $N(0, \sigma^2, I)$.

As can be seen, Equation (1) is a standard linear model. It could be said that each context represents its own sample of interest. However, the point of hierarchical linear modeling is that there are phenomena associated with the context that determine the variability of the β_{jk}

across contexts. Therefore, this model deviates from the standard linear model in that the within-group regression coefficients are allowed to vary across contexts. The equations that specify the posited relationships between context and within-group regression coefficients view the β_{jk} as outcomes in the second stage model. Thus, for the k regression coefficients in (1):

$$\beta_{jk} = \gamma_{0k} + \gamma_{1k}Z_{1j} + \gamma_{2k}Z_{2j} + \dots + \gamma_{p-1k}Z_{p-1j} + U_{jk} \quad (2)$$

where,

U_{jk} represents random error in this context level equation;

Z_{pj} are values on the context level variables for context j ; for $p = 0 \dots P-1$ independent variables in the second stage model; and,

γ_{pk} are regression coefficients that capture the structural effects of macro-level variables on micro-level relationships, β_{jk} .

The assumptions associated with Equation (2) are

- iii. The values of Z_{jk} are fixed - estimates are conditional on the sampled values.
- iv. U_{jk} are approximately $N_p(0, \phi_{kk})$ and the $cov(u_{qk}, u_{rk}) = \phi_{qr}$ where ϕ_{kk} are macro-level error variances and ϕ_{qr} are macro-level error covariances, $q = 0, \dots, k$; $r = 0, \dots, k$; q not equal r .
- v. The micro-level errors are independent of the macro-level error; i.e., for all i, j , and k , R_{ij} is independent of U_{jk} .

The set of relations implied in Equations (1) and (2) with assumptions i. to v. characterize the general

multilevel perspective on the substantive realities of school or classroom research. These equations can be combined to yield a single equation for the within-group outcome variable Y_{ij} :

$$Y_{ij} = \gamma_{00} + \sum_{k=1} \gamma_{0k} X_{ijk} + \sum_{p=1} \gamma_{p0} Z_{pj} + \sum_{k=1} \sum_{p=1} \gamma_{pk} X_{ijk} Z_{pj} + \frac{U_{j0} + \sum_{k=1} U_{jk} X_{ijk} + R_{ij}}{\text{error term}} \quad (3)$$

The above model is a mixed model because there are fixed coefficients (the γ 's) and random coefficients (the u 's and r 's). Further, it is covariance component model because the random coefficients covary.

Raudenbush and Bryk (1986) demonstrated that this model can be used to achieve several important objectives:

1. the decomposition of any observed relationship into its within and between-group components. Estimation of both an average within group and between group regression equation is provided for.
2. a multivariate formulation for examining the effects of between-group characteristics on within-group relationships.
3. adjustment of the within-group regression coefficients, β_{jk} , for other confounding variables within groups.
4. weighting of the estimated slopes, β_{jk} , in proportion to their precision in the regression against group-level factors (a characteristic of the empirical Bayes estimation to be discussed later). Greater precision is also achieved by using information on the correlation among the within-group regression coefficients when estimating the γ_{jk} .
5. provision of better estimates for the within-group regression coefficients β_{jk} than are available

through a traditional regression model that only uses the data from group j (another advantage to the empirical Bayes estimation which "borrows strength" by using the full data)²

Estimation of the Hierarchical Linear Model. Overall, the HLM approach attempts to provide estimates of the parameters from models of the form of (1) to (3) by employing procedures that allow for random effects, mentioned above, in the study of contextual impact on individual behavior. Empirical Bayes methods provide a comprehensive approach to the estimation of (a) point estimates and confidence intervals for the γ 's and, (b) since the β 's are assumed random, expectations, variances and covariances among these components. In the interest of clarity, a simple case best illustrates the logic of this approach.

First, let us assume that the within-unit outcome variable is a function of a single predictor plus random error and that the data have been centered around the unit mean such that the intercept term β_{j0} is zero.

$$Y_{1j} = \beta_j X_{1j} + R_{1j} \quad (4)$$

It should be noted that this equation is merely a simplification of Equation (1); thus, all previously discussed assumptions apply. Second, in Equation (2), no knowledge of context level factors (Z) that influence β_j is assumed, such that the between-unit model becomes,

$$\beta_{jk} = \gamma_{0k} + U_{jk} \quad (5)$$

This model specifies that a unit's slope is a function of the overall slope among the population of units and a component unique to unit j . The variance of U_{jk} , ϕ represents the true parameter variance among the population of units.

However, estimating the between-unit parameters poses some difficulty as these outcome variables (β_{jk}) are not directly observed. Although standard methods such as ordinary least squares can be used to obtain them, these estimates, β_{jk}^* , are measured with sampling error, which depends largely on the amount of data available in each setting.

$$\beta_{jk}^* = \beta_{jk} + e_{jk} \quad (6)$$

Under ordinary least squares theory, the errors, e_{jk} , are distributed normally with mean 0 and variance, v_j , where,

$$v_j = \sigma^2 / \sum X^2 \quad (7)$$

Therefore, the total variance of the observed within-unit slopes has a component due to true parameter variability and a component due to sampling error. If v_j and ϕ , are assumed known, the Empirical Bayes minimum mean squared error point estimators for $\beta_{jk}(B_{jk})$ and $\gamma_{0k}(G_{0k})$ are

$$B_{jk} = c_j * OLS(\beta_{jk}) + (1 - c_j) * G_{01} \quad (8)$$

$$G_{01} = (c_j * OLS(\beta_{jk}) / \sum(c_j) \quad (9)$$

and

$$c_j = \phi / (\phi + v_j) \quad (10)$$

where $OLS(\beta_{jk})$ is the ordinary least squares estimator of β_{jk} . The weighting factor, c_j , can be viewed as a reliability coefficient, a ratio of the true parameter variance to the observed variance. Thus, the Empirical Bayes estimate will be close to the OLS estimate if there is little error variability while the EB estimate moves toward the overall slope estimate if the variability associated with a given slope is largely attributed to error variance.

In addition to minimizing the effects of the sampling variance through the use of the above weighting procedure, Raudenbush and Bryk (1989) give several other important properties associated with the estimates generated by this procedure. First, because the covariation among the coefficients is taken into account, the estimation procedures are fully multivariate. In other words, the more the parameters covary, the more precise the estimates. Second, this method of estimation allows for the distinction between true parameter variation and sampling variation. This partitioning of variance is of substantive importance as this knowledge allows the researcher to evaluate the adequacy of his model. Finally, it is possible with this

procedure to estimate the covariation among the parameters. This covariance can be of substantive interest in providing the basis of a maximum likelihood estimate of the correlation between the mean level of achievement (i.e., efficiency or "excellence") and the distribution of achievement within a school or classroom (i.e., equity).

The above discussion assumes that the variances, v_j and ϕ , are known. Such an assumption is an aid to understanding the logic of the estimation procedure but is seldom tenable in practice. In the past, the fact that these variances had to be estimated placed severe limitation on the application of HLM's. However, the development of the EM algorithm by Dempster, Laird, and Rubin (1977) "affords a theoretically satisfactory and computationally manageable approach to covariance component estimation in hierarchical linear models" (Raudenbush & Bryk, 1986, p. 6).

Applications of the Hierarchical Linear Model

Raudenbush and Bryk (1986) provide an excellent example of using HLM to study interactive contextual effects, which will be used in a later section to illustrate the construction of a typical HLM model. However, theirs is but one application of multilevel analysis. In this portion of the review, other possible applications of the hierarchical linear model will be offered.

Inquiries into Individual Growth. Historically, inadequacies in conceptualization, measurement and design

have plagued research on change. From a conceptual point of view, a model for any phenomenon under study is needed to guide inquiry into the phenomenon. Yet research on individual change rarely identifies an explicit model of individual growth. Regarding measurement, studies of change typically use instruments that were developed to discriminate among individuals at a fixed point in time. These types of measures are inadequate for distinguishing rates of change among individuals. Finally, and probably most important, is the problem of design. Most studies of change are based on two time points. Such designs are inadequate for studying individual change (Bryk & Raudenbush, 1987). Fortunately, the development of HLM offers a set of techniques for research on individual change.

The logic of inference developed in the previous section can be applied in a straightforward manner to the study of change. The time-series data can be seen as nested within each subject. Therefore, the within-group model becomes the within-individual model and represents individual growth for each subject. This arrangement allows the researcher to proceed without difficulty when the number and spacing of time points vary across subjects.

In the study of change, it is assumed that growth parameters (the within-subject regression coefficients) will vary across individuals so a between-individual model is

used to represent this variation. This between-individual model represents each subject's growth parameters as outcome variables to be explained by subject characteristics.

In their analysis of preschoolers in Head Start, Bryk and Raudenbush (1987) showed that HLM could be used in the study of individual change to:

1. describe the structure of the mean growth trajectory
2. estimate the extent of individual variation around mean growth
3. assess the reliability of measures for studying both status and change
4. estimate the correlation between entry status and rate of change
5. examine how background and instructional variables influence change

These long-standing difficulties in the measurement of change were profitably addressed by the modeling of hierarchy.

Inquiries into Aptitude by Treatment Interactions (ATI's). Considerable effort by psychologists has been given to identifying interactions between student aptitudes and the "treatments" to which they were exposed. The logic being, of course, that certain students learn better from some instructional practices than they do from others. However, despite its appeal substantively, the results of ATI work have been mixed, at best, when they exist at all.

Burstein, Miller, and Linn (1981) posit that there is potential gain in viewing ATI research from a multi-level

perspective. They suggest rather than using the conventional strategy of including interaction terms among the explanatory variables, each classroom, or instructional unit, could be viewed as a "treatment" whose characteristics can be measured along several dimensions and then modeled within a multilevel framework.

For example, if carried out as a contrast between highly structured and unstructured treatments, the relationships could be modeled much as Raudenbush and Bryk did with the Catholic and public sectors. According to Burstein, et al. (1989), however, no ATI analysis of this sort has been carried out.

Inquiries into Differential Learning Opportunities.

Hierarchical linear modelling could also be used in the study of differential learning opportunities in the classroom. The relationships of characteristics upon class entry (i.e., initial ability, prior performance, social and psychological predisposition toward course content) to performance following instruction could be examined. More specifically, how do the mechanisms that teachers employ to get their goals accomplished, (individualizing instruction or using instructional groups; varying the content, level of presentation, pacing, choice of content strategy, or choice of level at which instruction is targeted) that result in differential learning opportunities affect the relationships

between entry characteristics and post-instruction performance?

Using HLM to More Accurately Estimate School Effects.

Raudenbush and Bryk (1989) make a case for using multilevel analyses for more accurately ranking schools by their effect. This application is certainly the most timely given our present "high stakes" uses of achievement test data. These researchers argue that efforts to assess school or teacher effectiveness without specifying the particular features which produce effectiveness will generally yield untrustworthy results. This means, for instance, that procedures which rank schools by averaging residuals from multiple regression or by estimating school-specific intercepts after controlling for school intake variables, such as student background, cannot, in general, be trusted.

From this perspective, ranking only makes sense when the variation in student outcomes depends on variation in school policies and practices. These researchers argue that exclusion of relevant policy variables in equations of school effectiveness introduce considerable bias to rankings. The direction of the bias introduced by ignoring policy variables can favor schools which are either advantaged or disadvantaged on composition variables, such as SES. Interestingly, however, their experience suggests that most often schools with advantaged student bodies will appear less effective than they really are. In other words,

the relatively high mean achievement of advantaged schools will be attributed too much to the advantaged backgrounds of their students and too little to the effectiveness of the teachers and school policies. Multilevel analyses such as HLM can minimize such bias by controlling for relevant pre-existing differences among students and including policy variables that can help to identify "meritorious schools." The challenge comes in formulating an explicit model of school quality. Without such a model it becomes difficult to be certain that the effects of school composition have been disentangled from other school factors with which composition is often correlated.

An Advanced Application: The Three Level Model. It is theoretically possible to model infinite levels of hierarchy. However the estimation of parameters in an "infinite" model given present algorithms is quite another matter. Yet Raudenbush and Bryk (1989) have successfully combined the interactive contextual effects model and the individual change model so that the growth of an individual learner within the organizational context of the school can be studied. This combination has provided for a three-level model.

Not surprisingly, the resultant data yields a dense web of empirical information. The following is only a partial list of the kinds of information provided by this three-level model.

1. Structural effects at the individual and school level can be disentangled.
2. Variance/covariance can be partitioned into within and between school components. (This gives valuable information about the structural sources of variation which would be useful in formulating future models.)
3. Variance partitioning also provides evidence about the reliability of the data for measuring status, learning rates, and other effects at both the individual student and school mean level. These are useful in interpreting the results from the structural analyses. This partitioning can also illuminate instances where the proposed statistical test is incapable of distinguishing between the competing equations. (Low reliabilities)

Conducting an Analysis: An Illustrative Case³

For the purpose of understanding the steps involved in a multi-level analysis, it is informative to look at a "landmark" application of HLM. In 1982 Coleman, Hoffer, and Kilgore (1982) suggested that academic achievement was more equitably distributed in Catholic rather than public schools. Field research conducted by Bryk, Holland, Lee and Carriedo (1984) suggested that this more "even" distribution of achievement was an outgrowth of the academic organization and normative environment of these Catholic high schools rather than the "common school" effect suggested by Coleman, et al. (1982). Raudenbush and Bryk (1986) applied an HLM framework to the Coleman, et al. (1982) data in an effort to determine what effects the internal organization of schools has on the distribution of achievement.

The outcome variable in these analyses was a standardized mathematics achievement score. The social distribution of mathematics achievement in each school was represented by a within-model that regressed mathematics achievement on minority status (MNRTY80), social class (SES), and academic background (ACDBKGD):

$$MATH\ ACHIEVMT = \beta_0 + \beta_1 (MNRTY80) + \beta_2 (SES) + \beta_3 (ACDBKGD) \quad (11)$$

Therefore each school's distribution of achievement was characterized in terms of four parameters: an intercept and three regression coefficients. Raudenbush and Bryk (1986) chose to center the SES and ACDBKGD variables around their respective school means so that the four parameters could be interpreted as follows (Bryk & Raudenbush, 1989, p. 167):

- β_0 - "base" achievement in school j
- β_1 - minority gap in school j (the mean difference between the achievement of white and minority students)
- β_2 - the differentiating effect of social class in school j (the degree to which social class differences among students relate to senior year achievement)
- β_3 - the differentiating effect of academic background in school j (the degree to which differences in the academic background of students eventuates in senior year achievement differences)

It should be noted that a school effective in equalizing the distribution of achievement would have a high base level of achievement, a small minority gap, and weak differentiating effects of class and academic background.

The Unconditional Model. The first step in this analysis, and one which is the starting point for almost any application of HLM is the unconditional model. For each within unit regression coefficient, β_{jk} , the between-unit model is:

$$\beta_{jk} = \mu_k + U_{jk} \text{ for } k=0,1,2,3 \quad (12)$$

No knowledge of macro-level factors that influence β_{jk} is assumed at this point. Thus, the within-group regression coefficient is seen to be a function of a grand mean plus random error. By estimating this "average" regression equation for the schools, two basic questions can be answered (Raudenbush & Bryk, 1988):

1. On average, is there a significant effect of each of the student background variables on math achievement within schools?
2. To what extent does each background effect vary from school to school?

The first question can be answered by testing the hypotheses that the gamma coefficients (which are synonymous with μ_k in Equation 5) are equal to zero. Under the null hypotheses, the γ_{pk} have asymptotic z distributions. In their study, Raudenbush and Bryk (1986) found that all of the student background variables considered had significant effects on math achievement with schools.

Estimated parameter variances for each of the β_{jk} are also of interest because hypotheses about these parameters address the second question. Under the null hypotheses, $\text{var}(\beta_{jk}) = 0$, the test statistics have asymptotic chi-square distributions with $J-1$ degrees of freedom (Hedges, 1982). Raudenbush and Bryk (1986) concluded that the minority gap, the social-class differentiation effect, and the academic background differentiation effect each varied significantly from school to school.

The unconditional model not only provides an estimate of the mean regression equation for the entire sample of schools, but it also provides estimates of the total parameter variances and covariances among the random effects. Expressed as correlations, these estimates describe the general structure among the distributive school effects.

Lastly, the unconditional model gives an indication of the reliabilities of the random effects. The reliability information is important because it provides the investigator with a sense of how much of the variability among a set of regression coefficients is likely to be explainable by school or class characteristics. Because parameter variance (i.e., the proportion of the observed variance expressed as the reliability) is the only potentially explainable variance component, failing to decompose the observed variance into its component parts

could lead an analyst to conclude that a model was inadequate even if it accounts for most of the explainable variance.

To summarize, the unconditional model is the first step in applying the hierarchical linear model. It provides a mean regression equation which allows the researcher to test each of his fixed regressors and the variability of the structural relationships across schools or classrooms. This stage of the model also provides correlations between the structural relationships. Further, the model enables the researcher to decompose the observed variance into its parameter and sampling components and express these as proportions in the form of reliabilities.

Next Steps: Proceeding with the Analysis. Although the unconditional model is common to all applications of HLM, further development of the model must be based on the individual researcher's purposes. In the case of Raudenbush and Bryk (1986), having concluded from the unconditional model that mean achievement as well as the structural relationships within schools differed across schools, and before investigating their main variables of interest - academic organization and normative environment, the researchers attempted to account for these differences (the reported "common school" effect posited by Coleman, Hoffer, and Kilgore (1982)) by introducing into the model what they termed "compositional" and "contextual" effects.

According to Bryk and Raudenbush (1989), a compositional effect is "the influence a school's social class, academic background, and minority concentration have on individual achievement" (p. 170). A contextual effect is "represented by including the school aggregate of a student-level variable in the between-school model for a slope coefficient" (p. 170). In short, a compositional effect is an aggregate variable included in the "intercept" (in the present example, base achievement) between-unit equation and a contextual effect is an aggregate variable included in the other between-unit equations. In general, the results of a compositional and contextual effects model failed to explain away the "common school" effect.

As a consequence of these results, the final step in the HLM application by Raudenbush and Bryk (1986) involved modeling the distribution of mathematics achievement as a function of characteristics of the academic organization and normative environment of schools. This model sought to answer three questions:

1. Do these organizational/environmental variables account for parameter variance in the within-unit structural relationships?
2. Does the Catholic advantage still persist once these variables are added to the model? (If these variables are important to explaining differences in the distribution of achievement, then sector effects in the former model will disappear.)
3. After modeling each within-group regression coefficient as a function of school level organizational and environmental variables, is

there evidence of significant residual parameter variation that remains unaccounted?

In order to develop the explanatory model, the researchers began with the context effects model described above.

The sector effect on base achievement disappeared when the average number of math courses taken by students, the school average hours per week students spend on homework, and a measure of staff problems were entered into the model. As might be expected, Raudenbush and Bryk (1986) found that achievement is higher in schools where students take more math courses, where they spend more time on homework and where staff problems are fewer. In addition, greater variability in math course taking and larger school size were both associated with a more unequalizing distribution of achievement in school both in terms of social class (SES) and academic background. Moreover, schools in which discipline is rated fair and effective by students are less differentiating on the basis of SES and academic background.

The pattern of effects for staff problems found by Raudenbush and Bryk (1986) begs comment on both conceptual and methodological grounds. In this study, it was found that staff problems had inverse relationships with the structural relationships which deal with differentiation of achievement based on social class and academic background. That is, schools with a large number of staff problems were less differentiating on the bases of SES and academic background. This may seem counterintuitive. However, what

this relationship says is that schools with staff difficulties reduce all student achievement to the lowest common denominator - everyone does poorly! Conceptually, such a distribution of achievement is certainly "equitable." However, this finding highlights the importance in school and teacher effectiveness research of balancing the notion of equity with the equally important concepts of efficiency and quality.

Methodologically, this finding also illustrates an interesting aspect of HLM that comes from the doubly multivariate structure of the between group model - multiple independent variables for multiple outcomes with a full covariance matrix. School characteristics may produce a web of interrelationships with the random effects. Therefore, users of HLM must be careful in the specification and interpretation of their models so that important observations like this one are not overlooked or misinterpreted.

Overall Raudenbush and Bryk's (1986) results provide strong evidence that academic organization plays an important role in changing initial differences in social class and academic background into differences in achievement.

Methodological Concerns: A Cautionary Note

From the preceding portions of this chapter, it is plain that much can be learned in the application of the

hierarchical linear model. However, multilevel models are still in their "statistical infancy," as it were, and must not be viewed as a panacea in the study of schools, teachers and their effects.

HLM has shortcomings, as does any statistical model. First, it is known that the empirical Bayes estimation procedures are not robust to non-normal data. When dealing with such data, the estimates tend to be "over-shrunk," perhaps masking relationships that do exist. Second, little is known about how various estimation procedures compare in terms of results. Therefore, explorations are needed of the varying results of the presently available estimation procedures because their properties may be dependent on the size and nature of the study population. Finally, because the models are so extensive due to the numbers and types of parameters to be estimated, there may be tradeoffs between conceptual appeal and the robustness of the corresponding estimation procedures. To obtain the advantages of HLM, the researcher must be very parsimonious in choosing the set of microlevel effects to model. Otherwise, costs of estimation explode at the same time that the quality of resulting estimates, in terms of their precision and interpretability, erode.

The aforementioned difficulties associated with HLM assume that the researcher has data to analyze and that this data was produced using perfect measures. In reality, the

cost of gathering the necessary data for these models is expensive in terms of time and money. Consequently, much of the work done with these models has used archival data for their analysis (Aitkin, Anderson & Hinde, 1981; Bryk & Raudenbush, 1987; Kennedy, 1990; Lee, 1986; Raudenbush & Bryk, 1986).

With respect to "perfect measurement," in educational research, observed measurements are invariably used as proxies for underlying constructs of interest. These observable indicators may diverge from their desired constructs. Therefore some spurious results may occur due to invalidity or unreliability of instruments utilized. Bayes estimation deals with such measurement problems indirectly by "shrinking" estimates back toward some central tendency or by giving less weight to extreme cases. However, there is as yet no way to model hierarchical data and use multiple indicators. Table 2.1 summarizes these and other methodological concerns and their associated problems which are addressed by multilevel models, specifically HLM.

Although these issues are problematic, the most serious difficulty facing the use of these models is lack of theory on what it is exactly that makes schools and teachers effective. In order for these mathematical models to find a larger audience than the statistical community, their development must be accomplished through real observations

in real classrooms for real students. It is to this end that the present study addresses itself.

Table 2.1

Methodological concerns, problems associated with these concerns and how these are addressed by HLM.

CONCERN	HLM SOLUTION
Small sample sizes a. bivariate regressions lack precision b. difficult to detect slope heterogeneity c. statistical power may be less	Data pooling in REML/Bayes estimation lessens the impact of small sample sizes
Ill-conditioned data a. outliers can dominate slope estimation within groups b. asymmetrical distributions c. nonlinearities	Data pooling in REML/Bayes estimation makes parameter estimates less dependent on outliers
Dependence among Observations within groups a. errors in same class are correlated b. variance of errors could fluctuate as a result of unequal-sized classes	Complex error structure is directly incorporated into the estimation process.
Non-random group composition can lead to both substantive and spurious (artifacts of selection) effects of macro-level properties	HLM does not directly respond to this problem.
Fallible Measurements a. validity - may measure more than one construct b. reliability - unsystematic fluctuation over occasions, observers, etc.	HLM offers no special protection against measurement problems

Notes to Chapter Two

¹The notation and assumptions for the model have been taken primarily from Bryk & Raudenbush, 1989 and Burstein, Kim, & Delandshere, 1989.

²These objectives have been paraphrased from a list provided by Raudenbush & Bryk, 1989.

³The study used as an illustration is the reanalysis of the High School and Beyond Data performed by Raudenbush & Bryk (1986). All results and interpretations are taken directly from this work.

CHAPTER THREE

METHODS AND PROCEDURES

Overview

The current chapter provides information about data collection methods and statistical procedures involved in the present study. It begins with a description of the population from which the subjects were sampled and the sampling procedures utilized. A discussion of the variables (and their operational definitions) is then pursued. Finally, the statistical models and their incumbent analysis strategies are delineated.

Population and Sample

In the current study, the population of interest consists of public primary school students in Louisiana. In an effort to increase the generalizability of the study to the entire state, twelve schools were chosen from two districts - a large southeastern city and a rural area near that city. These districts were chosen for their representativeness as well as their experimental accessibility.

For purposes of addressing the central questions of the study, it was necessary to choose schools within these districts which displayed a mixture of low and middle socioeconomic status (SES). This was accomplished through

the use of frequency counts of children participating in the free lunch program and those not participating. Only those schools which had more than 30% "free lunch" students and less than 70% "free lunch" students were retained in the sample, thus eliminating those schools with predominantly high or low socioeconomic compositions. It is important to note that it is the policy of both districts involved in the study to assign students to classes in such a way that ratios of low to middle SES within classrooms reflect overall school SES ratios. Therefore, classes included in the final sample reflect a variety of socioeconomic status.

In general, this sample of socioeconomically mixed schools was then stratified into two groups - effective and ineffective - through a multiple regression procedure. From these two strata, a random selection of six schools was made. Before discussing the specifics of the multiple regression procedure, a brief digression is in order to clarify the issues involved in classifying schools for this project.

There has been much debate in recent literature about the efficacy of using ordinary least squares residuals to estimate school effectiveness. Aitkin and Longford (1986) have noted that such residuals, which ignore students' membership in schools can be quite misleading. Bryk and Raudenbush (1992) suggest that empirical Bayes residuals estimated using HLM provide relatively stable estimates when

school sample sizes are small and which take into account group membership even when the number of groups is large. Efron and Morris' (1975) and Morris' (1983) reviews of both theoretical and empirical evidence indicate that these EB residuals will be a more accurate estimator of school effects due to a smaller mean squared error than the OLS school mean residual arrived at through the traditional multiple regression procedure.

However, it was deemed that traditional multiple regression was sufficient for the purposes of the current project for several reasons. First, in order to formulate school estimates using HLM, data at the student level is required. Such data were not available to researchers involved in the present study. Second, despite the technical advantages of EB estimators, there remain unresolved validity issues associated with these statistics as indicators of school performance. Among these are bias (discussed in Chapter Two) and "shrinkage as a self-fulfilling prophecy" (Bryk & Raudenbush, 1992). In terms of shrinkage, Bryk and Raudenbush (1992) point out that the estimators from the HLM model are conditionally biased. In other words, it is their nature to pull estimates of school effect toward the predicted value based on student background, to the extent that the OLS estimate is unreliable. "This procedure...operates as a kind of statistical self-fulfilling prophecy in which, to the extent

the data are unreliable, schools effects are made to conform more to expectations than they do in actuality" (Bryk & Raudenbush, 1992, p. 129). Finally, and most importantly, these estimates of school effects have not been found to differ significantly from traditional regression residuals results particularly at the extremes of distributions - which are the focus of the present study (Fitz-Gibbon, 1991; Kennedy, Teddlie, & Stringfield, 1991; Tate, 1988).

Accordingly, a forward stepwise regression procedure which identifies the smallest set of variables that are needed to maximize the explained variance was utilized to predict student achievement from various indices of socioeconomic status - free lunch status, parent's education, parent's employment, and interaction terms among these variables. Student achievement was calculated as Z-scores, using state means and standard deviations for each subject area and grade level of the state criterion-referenced achievement test. Then an overall school mean Z-score was computed.

The stepwise regression was carried out mechanically. The predictor variable accounting for the most variance in the achievement data was chosen first. Then, one at a time, other variables, or interactions, were added which account for the most remaining variance. This process was continued until the increase in the explained variance by adding another variable was insignificant. The use of the forward

stepwise regression was justified in this instance as there was no foreknowledge of which SES variables were the best predictors of achievement for this data.

As a result of the regression procedure, it was found that free lunch status explained 45% of the variability in school achievement. The data on parent's occupation, education, or interaction terms among these variables and free lunch status added very little to the explained variance.

Because regression residuals can be used as indicators of school effect, after the influence of socioeconomic status is partialled out, such residuals for each school were calculated in order to place them within the effective or ineffective category using the above model. A positive residual would indicate that the school's achievement was higher than predicted while a negative residual would indicate the school's achievement was less than predicted. Thus, a positive residual would result in a school being classified as effective. It should be noted that this procedure, along with the stepwise regression, is a practice often used in school effects studies (e.g., Brookover, 1979; Teddlie, et al., 1984)

Studies in the past have used the ± 1 standard deviation as the cut off for classifying schools as effective/ ineffective. However, Lang (1991) found that such a cut off point was not reliable due to the incidence

of a number of false positives and recommends $\pm .674$ as being the most reliable decision point. The distribution of the data for the present study was such that a ± 1 cut off was too limiting so a decision was made to use $\pm .75$, a decision point closer to Lang's (1991) recommendation. Therefore, schools with residuals of over $+.75$ standard deviations above the mean of the residuals were labeled "effective" and those with residuals under $-.75$ were considered "ineffective".

After the random selection of six schools from each strata was made, five teachers from each school were randomly selected, yielding a total sample of 60 teachers. Only teachers from third and fifth grades were considered as these were the only grades for which state criterion-referenced achievement test (LEAP test) data are available. Students in these teachers' classrooms were sampled exhaustively.

Variables

The HLM strategy for analyzing multilevel data followed in this study (Raudenbush & Bryk, 1989) initially entails fitting a model of within-class processes. This model reflects the social distribution of achievement within each classroom by linking socioeconomic status with achievement. It was expected that these structural relationships would vary across classrooms.

In the next steps, several between class models were built and tested which sought to explain the variability in the social distribution of achievement as a function of class (teacher) and school level variables. The between class models specified in the present study were formulated from those suggested by current literature (Emans & Milburn, 1989; Virgilio, 1987) and within the constraints of available data.

Operational Definitions

In this section the operational definitions of the variables used in this study will be presented. The operationalization of the school and class characteristics associated with the between class (second-stage) models are presented first, as these are considered as independent variables. This is followed by the definitions of the criteria used to formulate the link of SES to student achievement (the first-stage model). The SES/achievement relationship is used as the dependent variable in all between class models.

Variables for the Between-Class Models

School Effectiveness. Effective schools are defined as those whose students exhibit achievement at a level higher than that predicted from their socioeconomic data as measured by a predetermined residual cut-off score. This variable is obtained from the classification procedures described in the above section on sampling. To recap,

schools whose residual was at least $+.75$ standard deviations above the mean of residuals were labeled as "effective" and coded as "1". All other schools were coded as "0".

Time on Task. "Time on task" represents the amount of time students are actually engaged in a learning experience. A measure of this variable was obtained through the use of the Classroom Snapshot portion of the Stallings Observation System (Stallings & Kaskowitz, 1974) (Appendix A). Stallings and Kaskowitz (1974) conducted reliability studies of the instrument and reported inter-rater reliability of $.70$ on most variables. Stallings (1980) further found that this instrument was an effective predictor of student achievement. Inter-rater reliability was established for the three observers involved in the present study who used this instrument. The correlation was found to be $r=.93$

Teachers were scored on this variable on three occasions by three different observers. These observations were each one hour in duration. Observers were to make counts of students engaged in either interactions with the teacher or other adults present (i.e., an aide or volunteer), independent work or off task behavior and to make note of whether these interchanges took place in large or small groups or alone. The observers were instructed to begin scoring the Snapshot three minutes after the beginning of the class as designated by the school. Thereafter, observations were made at 7 minute intervals for 6

timeframes. The observers agreed prior to the visits that tallies would first be made of teacher interactions, then other adults, then students with other students or alone.

Scores for each teacher were obtained by calculating the number of students involved in tasks designated as "interactive" by the Classroom Snapshot (see Appendix A) and dividing this number by the total number of students in the class. This procedure was repeated for each of the 6 timeframes. The average timeframe percentage was then calculated.

Teacher Effectiveness. A measure of this variable was obtained through the use of the Virgilio Teacher Behavior Instrument (Appendix B). Teachers were scored on this instrument during the same visits discussed above. The development of this instrument was motivated by the need for an easily coded, research oriented instrument to assess teacher effectiveness in the classroom. Teddlie, Virgilio, and Oescher (1989) report estimates of internal consistency (Cronbach's alpha) as .96 for the total inventory and .88, .96 and .85 for the classroom management, instruction, and classroom climate subscales respectively.

Teddlie, et al. (1989) offer three ways of scoring the VTBI. First, a total mean item score may be used as an overall measure of teacher effectiveness. Second, mean item scores for three hypothetical scales (classroom management, instruction, and classroom climate) may be used. Third,

they state that users would be justified in reporting scores from the five empirically derived factor scores. The five factor solution yields one classroom management score, two distinct classroom climate scores (one which targets the physical classroom environment and one which focuses more closely on the emotional environment created by the teacher), and two instructional scores (one with a focus on delivery of material and questioning and one with a focus on monitoring and opportunities to learn). As it is the purpose of the present study to identify more specific categories of teacher behavior that impact the SES/achievement relationship within classrooms, the five score scheme will be used in the between-class models.

Variables for the Within-Class Model

Socioeconomic Status. Socioeconomic status is defined with respect to responses given by students to a survey questionnaire (Appendix C) which asked for specific information about parent education level and parent occupation. The scores for parent education level range from 1 (finished middle school) to 6 (went to graduate school after college). The parent occupation questions were open-ended to provide something of a "check" on the responses to parental education level. To further ensure student accuracy in reporting the data, researchers involved in the study were available to provide assistance to the children as the questionnaires were completed. In addition,

student responses were verified through the classroom teacher. Because it has been identified by several researchers (Fitz-Gibbon, 1991; A. Tashakkori, personal communication, August 26, 1993) as being a more reliable predictor of student achievement for similar data in Louisiana and England, father's educational level will be used as the indicator of socioeconomic status.

Student Achievement. For the present study, the criterion-referenced test results for grades three and five will constitute the student achievement measures studied. These tests are administered annually to grades three, five, seven, ten and eleven as part of the Louisiana Educational Assessment Program (LEAP). The tests for grades three and five are designed to give a measure of how well individual students, schools, districts, and the state have addressed the grade-level curricula in language arts and mathematics (Louisiana Department of Education, 1989). The items on the test are designed to reflect the specific standards of the state's curriculum guides.

Measures for internal consistency for the third grade test were reported as .93 for the mathematics portion and .94 for the language arts portion. Grade five reliabilities were .93 for both portions of the test. These levels of consistency were calculated from the actual 1989 LEAP administration using the KR-20 measure of reliability (Louisiana Department of Education, 1989). Content

validity of the LEAP was established in the development phase of the item bank which was designed to provide items which were matched on both content and item difficulty.

In order to have a single index of achievement for each student, the raw scores for both grade levels and subject areas were converted to T-scores using state means and standard deviations for the appropriate grade and subject. A mean score for each student was then calculated and used as the dependent variable in the within-class model.

Statistical Model and Analysis Strategy

In the current project, HLM analyses was conducted with a computer program developed by Bryk, Raudenbush, Seltzer, and Congdon (1986). A preliminary operation to the analysis is to simply partition the total variance in achievement into its within- and between-classroom components. This information is helpful in assessing the explanatory power of subsequent models for variation at each level of aggregation. A very simple HLM model is required to estimate these variance components. Within classrooms, let Y_{ij} represent the achievement score for student i in class j , and let this outcome simply vary around the class mean achievement, β_{j0} , with variance σ^2 :

$$Y_{ij} = \beta_{j0} + R_{ij}, \quad R_{ij} \sim N(0, \sigma^2) \quad (13)$$

Between classrooms, the class means vary around the grand achievement mean, γ_{00} with variance ϕ :

$$\beta_{j0} = \mu_j + U_j, \quad U_j \sim N(0, \phi) \quad (14)$$

Equations (13) and (14) represent a one-way random effects ANOVA model where classrooms are a random factor with varying numbers of students in each class sample. Using the maximum likelihood estimates of the within- and between-class variances generated by this model, an estimate of the intraclass correlation can be computed.

Once the partitioning has been accomplished, the first step of the HLM program involves the computation of the actual within-class parameter estimates. The analyses for the present study involves one first-stage model. This model has SES regressed on achievement scores (LEAP test). This within-class model represents the social distribution of achievement within classrooms and provides estimates of two parameters for each class, each of which was adjusted for all other independent variables in the model: 1) mean achievement (call it "Base"); and 2) a regression slope of SES on achievement (Slope 1). Put simply:

$$Y_{1j} = \beta_{j0} + \beta_{j1} \quad (15)$$

$$\text{LEAP SCORE} = \text{MEAN CLASS ACHIEVMT} + \text{SES}$$

To ease interpretation and computation, SES was centered around its class mean. The centering technique results in the intercept, β_{j0} , representing class mean achievement for group j . "Group mean centering" (Bryk & Raudenbush, 1992) was considered sufficient for this

application of HLM as the simple achievement means irrespective of SES were of interest. Had "grand mean centering" (Bryk & Raudenbush, 1992) been used wherein individual predictors are centered around the grand mean, the intercept would have to be interpreted as the class mean achievement for group j adjusted for the effects of SES. It was believed that such an adjustment would not offer insights essential to the central questions of the study.

In the second step of the analysis an unconditional between-class model is fit to each of the estimated parameters in the within-school model. The model is considered unconditional because it includes terms for a grand average and random error only. This model will allow Research Questions (1) and (2) to be answered.

The computer program generates t statistics for each of the parameter estimates indicating whether its value is significantly different from zero. In addition, large sample chi-square tests are performed on the components of the variance-covariance matrix for this model. These tests indicate whether or not there is sufficient variability among the within-class parameters to proceed with the analysis. Finally, the residuals associated with the estimates of the within-class parameters are output to a separate data set and the normality assumption is checked. Further, a plot of these residuals is made against

predictors from the between-class model to determine if curvilinear or other relationships are present.

Finally, in the third step, the explanatory between-class variables (effectiveness sector, five scores from the Virgilio Teacher Behavior Inventory, and time-on-task scores from the Classroom Snapshot) are introduced in a new between-class model. In order to estimate the between-class parameters, the within-class regression coefficients are used as dependent measures in the new model. Results from this analysis will answer Research Questions (3) through (5).

There are two essential components to the computer output at this juncture: (1) a table of gamma estimates (the final between-class parameters) with their standard errors, t test statistics indicating whether the value is significantly different from zero; and (2) a table of estimated parameter variances for each of the within-class output variables (an intercept and two slopes), along with their degrees of freedom, chi-square test statistic, and significance level of each variance. By comparing these estimated parameter variances for the within-class outputs in the various between-class models, it is possible to obtain an index of the effectiveness of the between model in explaining true parameter variability. These statistics are interpreted in much the same way as R^2 in simple regression analysis.

CHAPTER FOUR

RESULTS

Overview

This chapter presents the results of the study. Descriptive statistics for variables included in both the within-class and between-class models will be presented. Then results for the HLM analyses will be considered. First the unconditional models which address Research Questions 1 and 2 will be discussed. Finally, the process used to build the final explanatory between-class models will be described. These models address issues posed in Research Questions 3 through 5.

Descriptive Statistics

Variables of the Within-Class Model

Because these variables form the basis of what are considered dependent measures in the between-class model, the descriptive statistics associated with them will be discussed first. Table 4.1 presents univariate means, standard deviations and value ranges for father's education, the variable used to define student SES, and achievement, defined as scores on the LEAP test.

As can be seen, the mean for father's education, ($M=3.645$), indicates that the average educational level for parents of students in the study was greater than completion

of high school. Frequency counts made of responses to each category of education on the student questionnaire (see Appendix C) indicate that 40.5% of students reported that their father finished high school, 26.4% reported some college/attended business or trade school, and 19.2% reported completing college. Only 10.1% of students related that their father had less than a high school education. These figures are not surprising given the geographical area from which these data are drawn. The area's employment base consists largely of chemical and petroleum related industry which, without exception, requires a high school diploma, at a minimum. Further, most trades in these industries include addition vocational training necessary for the specialized nature of most plant work.

The achievement data come from the averaging of T-scores for the language arts and mathematics subscales of the LEAP test to arrive at a total battery score. The mean reflected in Table 4.1 is greater than the mean for the population of schools used to transform the original raw scores ($M=50$), suggesting that the sample of students retained for the study have greater achievement overall than students not included in the sample. Given, the SES data just reported, such a result could be expected.

The Pearson correlation between these two variables provides evidence that for the sample as a whole there

Table 4.1

Descriptive statistics for variables in the within-class model.

Variable	Mean	Standard Deviation	Range	
			Min	Max
Father's Education	3.645	1.055	1	6
Achievement	61.544	10.928	27	77.5

(N=1300)

Table 4.2

Descriptive statistics for variable in the between-class model.

Variable	Mean	Standard Deviation	Quartile	
			Q ₁	Q ₃
School Type	1.511	.501	1.000	2.000
Management	3.552	.819	3.000	4.143
Monitoring/ Opp. to Lrn	3.149	.960	2.477	3.909
Delivery/ Questioning	3.281	.918	2.625	4.000
Emotional Climate	3.629	.853	3.000	4.200
Physical Climate	3.236	.914	2.667	4.000
Interactive TOT	.520	.226	.357	.673
Total VTBI	3.363	.734	2.806	3.980
Achievement (aggregated to class level)	61.848	5.192	58.740	64.830

(N=180)

exists a significant positive relationship between father's level of education and achievement ($r=.343$, $p<.01$). It will be seen, however, that this relationship within classrooms is attenuated by particular between-class characteristics.

Variables of the Between-Class Model

Univariate means, standard deviations, and first and third quartiles for the between-class variables are presented in Table 4.2. All of the teacher behavior variables display generally large standard deviations, denoting the existence of considerable diversity in the levels of these behaviors across teachers. Although appearing large, the standard deviation for "School Type" ($sd=.501$) is commensurate with its dichotomous nature. The standard deviation for aggregated achievement ($sd=5.192$) also indicates considerable variation among class means. It will be seen that particular between-class predictors account for some of this variability.

Table 4.3 presents the Pearson correlations among the class level predictors. These relationships bear some discussion due to their direction and magnitude.

In terms of direction, it is immediately apparent that there are no negative values in the table. All but one of the teacher behavior variables (Interactive Time on Task) are positively and significantly associated with class mean achievement. This finding is in agreement with previous

process-product research on teacher effectiveness reviewed in Chapter Two.

The classification of a school as effective also has a significant positive relationship with class mean achievement, as evidenced by the correlation between school type and achievement ($r=.326$). This is certainly not surprising given the school effectiveness research cited in earlier chapters. Additionally, school type is also positively correlated with four of the five scale scores on the VTBI. These findings are consistent with the Teddlie, et al. (1989) results indicating that teachers in effective schools display significantly more evidence of effective teaching characteristics than do those in schools classified as ineffective.

The absence of a significant correlation between Interactive TOT and Achievement is interesting. This is a variable that has been found to be positively associated with achievement in previous studies (i.e., Stallings, Cory, Fairweather & Needels, 1977; Stallings & Kaskowitz, 1974). However, the HLM results presented next will explore the connection between this variable and the within-class SES/achievement link.

In addition to the direction of the relationships depicted in Table 4.3, attention must be given to their magnitude as well. In particular, it must be noted that most of the correlations between the five scale scores of

Table 4.3

Pearson correlations among class level variables.

	School Type	Class Manage	Monitor/ Opp.Lrn	Delivery/ Question	Emot. Clim.	Phys. Clim.	Inter. TOT	Total VTBI
Classroom Management	.326*							
Monitoring/ Opp. to Lrn	.234*	.669**						
Delivery/ Questioning	.334**	.704**	.861**					
Emotional Climate	.256*	.696**	.543**	.642**				
Physical Climate	.184	.422**	.585**	.522**	.407**			
Interactive TOT	.034	.174	.316**	.335**	.151	.120		
Total VTBI	.332**	.835**	.909**	.909**	.822**	.719**	.269*	
Achievement (aggregated to class level)	.299**	.340**	.282*	.332**	.282*	.278*	.086	.365**

(N=60)

*p<.05

**p<.01

the Virgilio Teacher Behavior Inventory (Class Management, Monitoring/Opportunity to Learn, Delivery/Questioning, Emotional and Physical Climate) are relatively large. These statistics raise a question of collinearity among the predictors. Such collinearity, if not addressed, will cause difficulties in obtaining estimates for the between class model. For example, larger standard errors may result which could lead to unreliability of parameter estimates.

However, at present there are no clear benchmarks given in the literature for determining what strength of relationship is problematic for between-unit predictors when using HLM. Therefore, several strategies from the context of multiple

regression were used in an effort to assess the degree of multicollinearity present among the variables.

First, the bivariate correlations from Table 4.3 were examined for coefficients of .8 or larger (Lewis-Beck, 1987). Only one such coefficient was found between the two instruction scores from the VTBI (Monitoring/Opportunity to Learn and Delivery/Questioning). Nevertheless, examination of only bivariate correlations fails to take into account the relationship of a variable with all other variables. Therefore, all between-class predictors with correlations of .5 or greater were regressed on one another and tolerances ($1-R^2$ for a variable with respect to all other regressor variables, (SAS, 1985)) were inspected. Tolerances of less than .15 (A. Tashakkori, personal communication, October 1, 1993) would indicate multicollinearity with other variables in the model. No variable exhibited a tolerance of less than .18.

The above is encouraging. However, in the absence of clear guidelines on this issue in the specification of the between-class model, two separate explanatory models will be presented. The first will assume multicollinearity among the five VTBI scores and will, therefore, combine them to obtain a total battery score. The total score will be used as a between class predictor along with Interactive Time on Task and School Type. The second model will explore the predictors in the conceptually logical composites suggested

by Teddlie, et al. (1989) using an empirical model building strategy recommended by Bryk and Raudenbush (1992). It is hoped that the second model will allow for a more specific investigation of those teacher behaviors that influence achievement and/or the SES/achievement link. Descriptive statistics and bivariate correlations among these composite scores are presented later in this chapter in conjunction with the discussion of the second exploratory model.

Results of the HLM Analyses

The HLM analyses proceeded in three steps as outlined in Chapter Three: 1) apportioning variation between and within classes; 2) assessing the homogeneity of regression hypotheses; and 3) assessing the effects of teacher behaviors on within-class mean achievement levels and SES/achievement relationships. At each step, the latter built upon the former.

Apportioning Variation: The Fully Unconditional Model

The analysis began with fitting the equivalent of a random-effects ANOVA model in order to determine the total variation in achievement test scores existing between- and within-classes. The model can be termed "fully unconditional" as there are no predictors specified at either the student or class level (see Equations 13 and 14, Chapter Three). Table 4.4 presents results of this model.

Table 4.4

Results of the fully unconditional model.

<i>Fixed Effect</i>	<i>Coefficient</i>	<i>se</i>	<i>t-ratio</i>	<i>p-value</i>
Average Class Mean, γ_{00}	61.890	.611	101.205	.000
<i>Variance</i>				
<i>Random Effect</i>	<i>Component</i>	<i>df</i>	χ^2	<i>p-value</i>
Variance among class means, u_{0j}	18.042	59	301.27	.000
Level-1 effect, r_{1j}	90.093			

The average class mean was estimated as 61.890. The pooled within-class or Level-1 effect, $\hat{\sigma}^2$, was 90.093. This figure represents the total Level-1 variance. As will be seen below, some of the variance is explained as SES is introduced into the within-class model. The variance among the J class means, $\hat{\phi}$, was 18.042. With the information on the variances, an estimate of the intraclass correlation can be computed using the following:

$$\rho = \hat{\phi} / (\hat{\phi} + \hat{\sigma}^2) \quad (16)$$

This statistic measures the proportion of the variance in achievement scores that is between classes. The intraclass correlation for these data was estimated as 0.167, indicating that 16.7% of the total variance in achievement is located between classes. Conversely, 83.3% of the

variance in the achievement measure is potentially explainable by within-class factors.

Testing Homogeneity of Regression: The Unconditional Between-Class Model

The next model regresses the SES indicator, father's education, on achievement test scores (see Equation 15, Chapter 2) at the student level but includes no predictors at the class level. Specifically, it tests the null hypotheses that the mean SES-Achievement slope is zero (Research Question 1) and that the intercepts (mean achievement) and slopes do not vary across classrooms (Research Question 2) by providing the average regression equation for the sample and estimates of the variances of the random effects. Also, residuals of the within-class parameter estimates are output to a separate data set and the normality assumption is checked through inspection of separate probability plots for each unit (Bryk & Raudenbush, 1992). No serious departure from normality was detected for the present data set. Finally, the homogeneity of variance for the within-class errors is also tested through the use of a likelihood-ratio test. For this data, the H statistic was 142.46 with 59 df ($p < .001$). This result indicates that heterogeneity of the Level-1 variance exists among the 60 classes in the sample. This may indicate that the Level-1 model has been misspecified perhaps because one or more important predictor variables may have been omitted from the

model (i.e., previous achievement). However, estimation of the class-level coefficients and their standard errors are rather robust to a violation of this assumption (Bryk & Raudenbush, 1992).

The results of this analysis are presented in Table 4.5. The average of the class means is estimated to be 61.889 with standard error .612 and t-ratio of 101.199 - The average SES/achievement slope is estimated to be 3.383 with standard error .288 and t-ratio of 11.759. This provides evidence that on average, student SES is positively, and significantly, related to achievement within classrooms, thus allowing the null hypothesis of Research Question 1 to be rejected ($p < .000$).

Table 4.5 also presents information regarding variability among the regression equations. Specifically, the estimated variance of the means ($\hat{\sigma}_{00} = 18.739$, $df=59$) gives clear evidence that highly significant differences exist among the 60 class means. Further, from the estimated variance of the slopes ($\hat{\sigma}_{11} = 1.280$, $df=59$) it can be inferred that the relationship between SES and achievement does indeed vary significantly across the population of classes. Hence, the null hypotheses of Research Question 2 may also be rejected ($p < .040$).

As the above chi-square tests are simple univariate tests that do not take into account the other random effects in the model, these results were cross-checked through the

use of a multivariate likelihood-ratio test which uses all of the data available (Bryk & Raudenbush, 1992).

Specifically, the deviance statistic for this model was compared to the deviance statistic for a restricted model with only a random intercept. For this data, the deviance statistic for the present model was 9,580.91 with 4 df. For the restricted model, deviance was calculated as 9,402.99 with 2 df. Hence the test statistic, H , was calculated to be 177.92 with 2 df - confirming the conclusion that classes do vary in their distributive effects ($p < .000$).

Additionally, information about the reliabilities of the random effects in the model are reported in Table 4.5. These reliability indicators were derived from the ratio of estimated parameter variance in each regression coefficient to the total observed variance in the corresponding estimated OLS slope. Results indicate that the intercepts are reliable, with approximately 83% of the variation in achievement potentially explainable by class level predictors. As expected, the regression coefficient associated with SES is less reliable indicating that 76% of the variation is attributable to sampling variance and not explainable by class level predictors.

It should be noted that the estimated Level-1 effect is now 76.746, a reduction from the fully unconditional model. An index of the variance explained by the present model was calculated by comparing the variance estimates from these

Table 4.5

Results of the unconditional between class model.

<i>Fixed Effect</i>	<i>Coefficient</i>	<i>se</i>	<i>t-ratio</i>	<i>p-value</i>
Overall Mean Achievement, γ_{00}	61.889	.612	101.199	.000
Mean SES/Achmt Slope, γ_{01}	3.383	.288	11.759	.000
<i>Variance</i>				
<i>Random Effect</i>	<i>Component</i>	<i>df</i>	χ^2	<i>p-value</i>
Class Mean, u_{0j}	18.739	59	353.67	.000
SES/Achmt Slope, u_{1j}	1.280	59	79.33	.040
Level-1 effect, r_{1j}	76.746			
<i>Reliabilities of OLS Regression Coefficient Estimates</i>				
Class Mean Achievement	.834			
SES/Achievement Slope	.239			

two alternative models. By adding SES as a predictor of achievement, the within-class variance was reduced by 15%. Therefore, it can be concluded that SES accounts for about 15% of the student-level variance in achievement.

One final statistic generated by the unconditional between-class model bears mentioning - that is the correlation between the intercept and the SES/achievement slope. This correlation was found to be $-.113$ which suggests that classes with high achievement tend to have weaker SES/achievement relationships.

Explaining Heterogeneity of Regression: The Explanatory Models

Using the conclusions of analyses of class mean achievement and relationships between SES and student achievement as they varied by class, this multilevel analysis involved creating explanatory models to determine which teacher behavior factors contributed to achievement and changes in the strength of association between SES and achievement – the subjects broached in Research 3 through 5. Therefore, the unconditional between-class model was extended to include these factors. Two different between-class models were explored. The first assumed serious multicollinearity between predictors while the second attempted to investigate more specific teacher behaviors.

Explanatory Model I: The VTBI Between Class Model. As explained in previous paragraphs, due to the relatively large correlations among five of the between-class predictors, a model using a total battery score for the Virgilio Teacher Behavior Inventory (VTBI) was examined. This model also included Interactive Time on Task (InterTOT) and School Type (SCHTYPE) as between-class predictors. The results from this explanatory model are presented in Table 4.6.

Table 4.6 shows that total score on the VTBI is positively related to achievement after adjusting for SES ($\gamma = 2.385$, $p < .011$). That is teachers who exhibit more

effective teaching practices globally engender higher achievement in their classes. Conversely, higher scores on the VTBI seem to be associated with more differentiation in terms of the distribution of achievement ($\gamma = 1.208$, $p < .007$). School type and Interactive Time on Task are not significantly associated with achievement ($t = .444$ and $.141$ respectively). However, with regard to the SES/achievement slopes, it appears that classes located in schools labeled as effective tend to have weaker SES/Achievement slopes than do classrooms in their ineffective counterparts ($\gamma = -1.349$, $p < .037$). Similarly, teachers who engage the highest percentage of students in interactive tasks have an ameliorating effect on the relationship between SES and achievement ($\gamma = -3.882$, $p < .003$).

Chi-square analyses were used to test hypotheses regarding whether residual differences among classes in a particular within-class parameter (either intercept or slope) can be attributed to nothing more than sampling variance. The χ^2 statistic of 286.36 ($p < .000$) indicates that after explaining variation in mean achievement by teacher variables, some significant variation remained. Conversely, the χ^2 tests of residual variance of the SES/achievement slopes indicates that after accounting for the effects of class-level predictors, no explainable variation remained ($\chi^2 = 63.079$, $p < .240$).

Table 4.6

Results from Explanatory Model I.

<i>Fixed Effect</i>	<i>Coefficient</i>	<i>se</i>	<i>t-ratio</i>	<i>p-value</i>
Model for Class Means				
INTERCEPT, γ_{00}	52.982	2.648	-----	-----
SCHTYPE, γ_{01}	.574	1.292	.444	.359
INTERTOT, γ_{02}	.361	2.557	.141	.393
VTBI, γ_{03}	2.385	.866	2.753	.011
Model for SES/Achievement Slopes				
INTERCEPT, γ_{10}	3.508	1.207	-----	-----
SCHTYPE, γ_{11}	-1.349	.610	-2.211	.037
INTERTOT, γ_{12}	-3.882	1.188	-3.267	.003
VTBI, γ_{13}	1.208	.412	2.930	.007
Variance				
<i>Random Effect</i>	<i>Component</i>	<i>df</i>	χ^2	<i>p-value</i>
Class Mean, u_{0j}	15.423	56	286.36	.000
SES/Achmt				
Slope, u_{1j}	.477	56	63.08	.240
Level-1 effect, r_{1j}	76.763			

Explanatory Model II: The Individual Teacher Behavior Model. Although the above model demonstrated significant relationships in expected directions, the original intent of the present study was to investigate teacher behaviors in a more specific manner (Research Question 5). Therefore, as has been stated, an empirical model building technique recommended by Bryk and Raudenbush (1992) was used in an effort to fulfill this aim.

These authors advocate the use of a simple univariate regression of the empirical Bayes residuals from each of the

J+1 equations on Z variables that might be added to the model. Approximate "t-to-enter" (Bryk & Raudenbush, 1992, p. 214) statistics may be computed in this way. The strategy then becomes an exercise in choosing the variable with the largest approximate t and entering it in the appropriate model, while monitoring standard errors as predictors enter. However, Bryk and Raudenbush (1992) caution that the statistics are only approximations as the target model is doubly multivariate due to multiple predictors and correlated errors across models. Nonetheless, "they will usually provide a good indication of the next single variable to enter one of the Level-2 equations" (p. 215). Given this caveat, and their additional recommendation to fit a tentative model for the intercept before fitting models for random slopes, the model for achievement was built before moving on to the model for the SES/achievement slopes.

It was expected that the relatively high correlations among the five scores from the VTBI would cause difficulties in finding predictors which made unique contributions to the explanatory power of the models. Thus, composites of the most highly correlated were created in an effort to alleviate this difficulty. Table 4.7 summarizes these consolidations. Specifically, the three score scheme for the VTBI cited by Teddlie, et.al (1989) was computed yielding a score for classroom management (MANAGE), instruction

(INSTRUCT), and classroom climate (CLIMATE) for each teacher. Descriptive statistics and correlations among

Table 4.7

Composites created from VTBI scores.

<i>Original</i>	<i>Composite</i>
Monitoring/Opportunity to Learn Delivery/Questioning	INSTRUCT
Emotional Climate Physical Climate	CLIMATE
Classroom Management	MANAGE

these predictors and achievement (aggregated to the class level) are presented in Tables 4.8 and 4.9, respectively. It should be noted that high correlations persist among these three composite scores, and any results derived from them should be viewed with caution.

Once these composites were constructed the process of building the model for achievement was begun. Examination of the approximate t -scores generated by the unconditional between-class model revealed that the composite MANAGE should be entered first into the intercept model. Once entered, however, no other composite reached significance. At this point, the process for fitting a model for the SES/achievement slopes was initiated. The largest t -to enter for the slopes model was Interactive Time on Task (INTERTOT), followed by INSTRUCT and School Type (SCHTYPE).

Beyond these predictors, no others reached significance. The outcomes of this exploratory model are offered in Table 4.10.

Table 4.8

Descriptive statistics for VTBI composite scores.

Variable	Mean	Standard Deviation	Quartile	
			Q ₁	Q ₃
INSTRUCT	3.207	.891	2.549	4.000
CLIMATE	3.424	.727	2.833	4.000
MANAGE	3.552	.819	3.000	4.143

Table 4.9

Pearson correlations among VTBI composite scores.

	INSTRUCT	CLIMATE	MANAGE
CLIMATE	.737*		
MANAGE	.716*	.667*	
Achievement (aggregated to class level)	.318*	.325*	.340*

(N=60)

*p<.01

The results displayed in this table are not profoundly different from those in Table 4.6. They have, however, been made slightly more specific. First, CLIMATE and INSTRUCT as composites distinct from the total VTBI score used in the earlier model were non-significant as predictors of achievement in the present model. Thus, it can be said that

teachers scoring higher on items which measure those skills involved in classroom management produce greater achievement within their classrooms ($\gamma = 2.387$, $p < .001$). Second, scores on instruction scales were significant in predicting SES/achievement slopes. However, classroom management was not important in predicting slopes within classrooms. So, it can be said, teachers with higher INSTRUCT scores tend to be more differentiating in the distribution of achievement in their classrooms ($\gamma = 1.029$, $p < .003$). The results for INTERTOT and SCHTYPE are identical to those specified in the earlier model.

Again chi-square tests were conducted on residual variances with similar results. After accounting for class-level predictors, significant variation remained in the intercept (achievement) while none remained in the slopes. This is not surprising recalling that the reliability of the slopes (.239) denoted that only some 24% of the variability was potentially explainable by class-level predictors.

The reduction in these variance components from the unconditional model is noteworthy. Table 4.11 summarizes the changes in estimated parameter variance for the random effects from the unconditional between class model to the present model. Differences between two estimates, expressed as a proportion reduction in parameter variance, $\%R^2$, relative to the unconditional model are also presented in Table 4.11.

Table 4.10

Results of the Explanatory Model II.

<i>Fixed Effect</i>	<i>Coefficient</i>	<i>se</i>	<i>t-ratio</i>	<i>p-value</i>
Model for Class Means				
INTERCEPT, γ_{00}	54.048	2.172	----	----
MANAGE, γ_{01}	2.242	.628	3.567	.001
Model for SES/Achievement Slopes				
INTERCEPT, γ_{10}	4.212	1.070	----	----
INTERTOT, γ_{11}	-3.961	1.153	-3.435	.002
INSTRUCT, γ_{12}	1.029	.319	3.229	.003
SCHTYPE, γ_{13}	-1.290	.577	-2.238	.035
Variance				
<i>Random Effect</i>	<i>Component</i>	<i>df</i>	χ^2	<i>p-value</i>
Class Mean, u_{0j}	14.907	58	286.20	.000
SES/Achmt				
Slope, u_{1j}	.303	56	61.466	.286
Level-1 effect, r_{1j}	76.824			

Table 4.11

Proportion of variance explained by models.

	<i>Class Mean Achievement (%R²)</i>		<i>SES/Achvmt Slopes (%R²)</i>	
Unconditional Model	18.739		1.280	
Explanatory Model I	15.423	(17.7%)	.477	(62.7%)
Explanatory Model II	14.907	(20.5%)	.303	(76.3%)

CHAPTER FIVE

DISCUSSION AND CONCLUSION

Discussion

These analyses have sought to identify teacher behaviors which have a positive effect on achievement while simultaneously ameliorating the effects of socioeconomic status on that achievement. The investigation focused on sixty classrooms - thirty from schools labeled as effective and thirty from schools labeled as ineffective. Unlike any previous study of teacher behavior, the present inquiry explored the effects of classroom practices on the distribution of achievement within the classroom as well as on mean levels of achievement through the use of the Hierarchical Linear Model (Raudenbush & Bryk, 1986). Additionally, unlike most studies using HLM, true behavioral data from the classrooms were collected for analysis.

Schools were labeled as effective or ineffective by using residual scores from a regression analysis. Those schools with achievement scores that were .75 standard deviations above what was predicted by the socioeconomic status of the school were considered effective while those with achievement scores a like distance below what was predicted were regarded as ineffective. Twelve schools, six from each effectiveness sector, were then randomly selected. From within these schools, five teachers from grades three and five were randomly selected.

Data on teacher behaviors were gathered through classroom observations during which six dimensions of effective teaching were evaluated. These behaviors were interactive time-on-task, classroom management, strategies for monitoring student progress and providing opportunities to learn, strategies for presentation of content and questioning techniques, social/psychological environment of the classroom, and physical attributes of the room. Not unexpectedly, these teacher behaviors were found to be positively and significantly correlated with one another.

Due to this finding and the lack of direction in the literature pertaining to the collinearity issue, model building proceeded from an exploratory standpoint. Once unconditional models were examined and their results indicated that there was significant variation in the class-level regressions, total battery scores from state achievement tests and the relationship between those scores and SES, measured by fathers' education, served as the dependent measures of two explanatory models. The first model dealt with the teacher behaviors in concert, while the second sought to isolate more specific teacher behaviors which might be associated with achievement and the relationship between SES and achievement in the classroom.

Impact of Class-Level Variables on Achievement

This project revealed that there was indeed variation at the class level with regard to achievement and that this diversity could potentially be explained by class level factors. For the present data, it was found that higher total scores on the Virgilio Teacher Behavior Inventory were positively related to the level of achievement in the classroom. Specifically, teachers who are proficient classroom managers; who are able to present relevant information clearly and to question effectively; who are diligent in monitoring student progress and ensuring the maximum opportunity for every student to learn; and who accomplish all of the above in an environment which combines high expectations for learning with a genuine investment of self in the students do increase the achievement of children in their charge. Undoubtedly, these findings confirm what has been shown with previous process-product research.

Yet, when the VTBI was broken into three operationally distinct scales, the only significant predictor was classroom management. Although it was expected that instructional concerns would be primary to determining the level of achievement in the classroom, perhaps this finding is not so surprising. It is clear that the teacher who cannot execute the mundane activities of a classroom, such as collecting lunch money, distributing materials, or gathering his/her students into a line for a trip to the

lunchroom or playground, in an efficient manner will most certainly not have time to instruct. This assumes, however, that given the time, the teacher knows how to use it effectively.

From a methodological standpoint, however, there may be another factor to consider. Even when aggregated into three composite scores, the VTBI scales were still relatively highly correlated. In view of the lack of information on the effects of multicollinearity at the group-level when using HLM, it is difficult to know what effects such correlations have on the estimates. Thus, it could be that one of the other scales may also impact achievement at the class level, but the model did not detect it.

The results concerning the relationships between interactive time on task, school type and class level achievement are incongruous with some previous research on teacher and school effectiveness. Neither of these variables was significant in predicting achievement at the class level.

With regard to the time on task variable, as suggested above, classroom management techniques may be preeminent when discussing the effectiveness issue. Perhaps Socrates, himself, would fail to impact achievement if he had difficulty prioritizing the bureaucratic tasks of his classroom, thus leaving little time for engaging his students in interactive tasks. In addition, some more

recent work in the area of time on task has shown that this variable may not be as important in impacting achievement as was previously suspected.

Teddlie, et al. (1989) demonstrated that teachers in effective schools exhibit more effective teaching behaviors than their counterparts in ineffective schools and that the variance of behaviors is smaller in effective schools. Therefore, logically, it would seem that the designation of a school as effective would denote more effective behaviors, ergo a significant impact on achievement. However, such was not the case with these data, although the relationship was positive in direction. It seems from these results that teacher behavior on a class by class basis has more to do with impacting achievement than the effectiveness or ineffectiveness of a school overall. Such a viewpoint makes sense when it is recognized that although there tend to be fewer ineffective teachers in effective schools with the opposite holding true in schools labeled as ineffective (Crone, 1992; Virgilio, et al., 1991), there are still teachers in each setting that do not fit their school profile. This finding points to the veracity of the notion that the keeper of the door to the classroom may truly be the determiner of student success.

It has been seen that while total VTBI and the composite score for classroom management significantly impacted achievement; interactive time on task and school

type did not. Of the variation in achievement potentially explainable by class-level variables, explanatory models I and II were successful in explaining approximately 18% and 20% of that variability, respectively. These percentages exceed those reported by Teddlie, et al. (1984) and Reynolds (1992). However, it must be acknowledged that after entering the teacher behaviors discussed above into the intercept models, there remained significant within-class variation. Clearly, factors not considered in this study are operating to a significant degree.

Impact of Class-Level Variables on the SES/Achievement Link

It should be noted at the outset that the present study's inclusion of an exploration of the within class SES/achievement relationship represents a unique contribution to the knowledge base of effective teaching, and the findings point to an intriguing configuration of results for class-level variables which impact this relationship. The pattern of direction differs somewhat from that anticipated at the study's inception. The questions posed in Chapter One concerning the within class slopes were directional in nature. It was believed that all of the class level factors included in the models would weaken the SES/achievement relationship. This belief obviously implies an equity perspective - one which focused exclusively on the academic progress of the lower end of the

socioeconomic continuum. However, as was seen in Chapter Four, this was not the case. The teacher behaviors measured with the VTBI showed the opposite trend, suggesting that particular actions on the part of teachers maximize the potential of all students, regardless of SES - a quality or an excellence focus. Although these results certainly warrant attention, the methodological concerns stated above for the intercept (achievement) models must be echoed for the SES/achievement models as well.

Interactive time on task was found to be inversely related to differentiation in the classroom on the basis of SES. Although studies in the past have been able to link this variable to mean achievement levels, the present study, for the first time, reveals that this variable is also instrumental in ameliorating the effects of SES on achievement. That is, academic activities which require direct contact with teachers, as opposed to independent seatwork-type assignments, help low-SES students to achieve at a level closer to that of their middle-SES classmates. From what is known about the lack of educational support in the homes of low-SES students, it makes sense that giving these students firsthand experiences with a supportive teacher would foster greater academic success more commensurate with middle-SES children in the classroom. Additionally, it could be said that these teachers are better at gaining and keeping the attention of all

socioeconomic groups in their classrooms. In this sense, they are less differentiating on the basis of SES.

Similarly, a negative relationship between school type and the SES/achievement link was also uncovered. Effective schools overall seem to render SES a less effective indicator of achievement within classrooms than do ineffective schools. Raudenbush and Bryk (1986) studied the differences between Catholic and public high schools and found Catholic schools to be less differentiating with respect to SES. Indeed, the majority of the literature does support such a finding. Placed in the framework of equity, this result is unquestionably in the expected direction.

Interestingly, however, the data reveal the opposite relationship when it is the classroom and not the school which is considered. A stronger association between socioeconomic status and achievement was uncovered in classrooms. The data establish that teachers who received higher scores on the total VTBI and the INSTRUCT composite tend to be more differentiating on the basis of SES. This finding seems to be at odds with the results for effective schools. Moreover, it is in direct opposition to the stance taken by equity advocates who assert that an effective teacher, or school, should not only raise overall achievement scores, but should also raise the scores of students from low socioeconomic backgrounds to reflect the performance of the entire group. These results, however,

are easily integrated with the previous finding of a positive association between the VTBI and mean achievement. Although low-SES students may be scoring below their middle-SES classmates in the classroom of an effective teacher, the entire group is outscoring students in the ineffective classroom. This substantiates Crone's (1992) findings at the class level. Teddlie, et al. (1984) and Mortimore and Sammons (1987) relate a similar result at the school level. All found that low-SES students in effective educational settings achieved at a significantly higher level than middle-SES students in ineffective settings.

The direct nature of the relationship between the SES/achievement link, total VTBI, and the INSTRUCT composite can also be viewed from the reverse perspective. It could be said that ineffective teachers are less differentiating with respect to SES because all of the students in their classrooms do poorly!

Considered separately, the findings for school type and the VTBI are not difficult to interpret. The obstacle comes when these two findings must be reconciled to one another. This seeming contradiction could be accounted for by the fact that only a few teachers in each school were part of the current study. Therefore, the possibility exists that although more differentiation is occurring in the included classrooms, for the school overall, the opposite trend is pervasive.

It is also possible that the definition of school type holds an explanation of this result. School type was determined by comparing predicted school achievement levels with actual achievement levels after controlling for percent free lunch. According to the HLM analyses, then, schools with higher than predicted achievement (based on student intake) are associated with flatter slopes, in other words, less association between SES and achievement school-wide. In these schools low SES students are performing above what might normally be expected of them. Inasmuch as this is true, these schools are less differentiating on the basis of SES.

Conversely, when examining behavior within the classroom, effective teachers in effective schools are maximizing the potential of all students. This maximization would certainly have the effect of raising mean achievement above what was predicted on the school level, but it also appears to have the effect of increasing the association between SES and achievement on the class level. Combined with the results for time on task, it is clear that effective teachers are getting and keeping the attention of all his/her students. Once they are attending to appropriate academic tasks, not only are low SES students achieving at a greater level than predicted, but also the cumulative effects of the more supportive higher SES home and effective teaching at school are propelling those students at the

higher end of the SES continuum above their predicted outcomes. Therefore, the association between SES and achievement is increased.

The observed inconsistency between class and school level results unquestionably suggests an area of research that must be explored. As Bryk and Raudenbush (1992) recommend, the models for classifying schools as effective/ineffective must be more detailed and must include more school policy-type variables. Had such variables been available and utilized for these data, the results for class and school levels may have been more compatible. These findings also indicate that modeling the relationship between socioeconomic status and achievement at only the level of the school as did Raudenbush and Bryk (1986) is insufficient for fully understanding those factors which mediate it. Other investigations similar to the present one are needed which utilize true behavioral data at the teacher level to investigate the effects of teacher action on the within class SES/achievement relationship.

Limitations of the Study

Limitations of the Model

It should be noted that the specification of the within-class model has a number of limitations. First, as the primary focus of the study is the examination of the relationship between SES and achievement, the reporting of

SES data by the children in the sample may be problematic. Research has indicated that elementary children may be incapable of accurately describing the educational level or occupation of their parents. Although these data were verified with classroom teachers, errors may still have occurred. Secondly, it is plain that father's education as a single within class predictor may have been inadequate. Certainly the inclusion of other contextual data would have allowed for a more accurate depiction of the distribution of achievement within the included classrooms. For example, modeling the relationships of variables such as percent minority, ethnicity, and gender with achievement would have all been informative. Additionally the interactions of these variables with SES may have proven significant. Investigation of the gender by SES interaction may have been particularly interesting given the current high level of concern for the susceptibility of young minority males to be victims or perpetrators of violent crime. Further, previous achievement data on the students would have enriched the results of the study. Although these data were too costly to collect within the confines of the current project in that such information was only available with the consent of every parent of the some 1300 students involved in the study, succeeding inquiries would benefit from their presence in both intercept and slope models.

Although all of the variables mentioned above are worthy of inclusion in subsequent investigations of the issues explored here, researchers must be cautious in the number of level-1 predictors which are specified as random when using HLM. The number of variance-covariance components to be estimated in a two level model rapidly increases with the number of random predictors in the level-1 model. As this number grows, significantly more information is required to obtain these estimates. It may be that particular models can sustain only a limited number of random effects (Bryk & Raudenbush, 1992). Nonetheless, the means and SES/achievement relationships depicted in the study were almost certainly mediated by one or all of the aforementioned variables and should be interpreted accordingly.

It should also be recalled that only two elementary grades were included in the sample. While pertinent to future academic success, there is evidence that the processes at work in the primary grades are distinct from those at the middle or high school level. Further, the schools from which the students were sampled were selected from an area to which travel was convenient. There were, however, both rural and metropolitan schools included in the sample. In addition, because the presence of the researchers required the approval of the principal of each school, their participation must be considered voluntary.

This suggests the possibility of structural differences in these settings and cautious generalization to non-volunteering schools is advised. Inasmuch as the achievement data were limited to only those students present on the days the test was given, a degree of unrepresentativeness is present in the data.

It must also be noted that the achievement data for this project were derived by combining scores on language arts and mathematics subscales of the LEAP test. Although this arrangement was adequate for the purposes of the present study, subsequent research may wish to build separate explanatory models for each of these areas. It is known that the processes involved in dealing with verbal and quantitative information differ. Therefore, it stands to reason that the relationship between SES and achievement in these domains may also differ. Investigations of these differences could be fruitful.

Limitations of the Methodology

Albeit useful for uncovering hertofore undiscussed interrelationships of variables like the within class SES/achievement relationship, HLM falls short of being the methodological savior of school and teacher effects. Specifically, the pattern of intercorrelation which was exhibited among the between class predictors is not an unusual occurrence, yet the effects of such multicollinearity have not been discussed by research

methodologists. As the use of multilevel models has quickly become the standard methodology for scrutinizing the effects of teachers and schools, studies which investigate the effect of such collinearity must be undertaken to better inform the population of users as to its consequences. Perhaps this could be done by examining the differences in parameter estimates and significance levels between coefficients generated using sufficient statistics matrices - the heart of any HLM analysis - built with variables with varying degrees of association.

In the same vein, information concerning the data used to build the vital sufficient statistics matrices is pointedly absent from studies using these models. Again, as the discussion and use of these models has spread from the methodological community to use by a more general practice oriented group of researchers (and thus perhaps less well versed in the methodological peculiarities of these models), this information is crucial for consumers of this research in determining if the models have been constructed from operationally distinct components.

Conclusion

The picture which emerges from these analyses is characterized by a conflict between the interests of "equity" and "excellence." On the one hand it seems that effective schools are more equitable in their distribution

of achievement, while on the other, effective teachers tend to be more likely to generate higher achievement for all even at the peril of equity, per se. Although most school and teacher effectiveness researchers would contend that these two concerns must both be served, it is unclear if they can be successfully dealt with concurrently given the present configuration of the public school in the United States.

It could be argued that equity in educational attainment is an unreasonable goal in that some person's or group's achievement must always be subordinate to that of some other person or group. If a teacher is truly effective for all the students in his/her class, they will all achieve at a greater rate. Thus, those who started at a disadvantage may appear to retain this disadvantage. However, as has been seen, such a judgement is completely dependent upon the group to which these students are compared.

Although equity for all students at all times is impossible, it should not be the case that the group which is consistently short-changed be those students with low-SES backgrounds. Research has pointed to the fact that low-SES children do not need the same kinds of instructional strategies as their more socioeconomically advantaged classmates. Perhaps it is time that these children were allowed the opportunity to be schooled in ways that meet

their special needs - even if this means giving these opportunities in a separate setting. It may be necessary for effective teaching researchers to investigate some of the schools now emerging which have chosen just such an approach.

Further, when considering the increased relationship between SES and achievement in the effective classroom, it would also be interesting to investigate the impact of a teacher's personal orientation with regard to the issues of equity and excellence on the SES/achievement relationship in his/her classroom. Various questions might be posed. Is the maximization of every child's potential - the increased association between SES/achievement - a conscious choice or a natural outgrowth of effective teaching behaviors? Does a teacher who embraces more of an equity perspective consciously target low SES children for additional attention, and what effect does this attention have on mean achievement levels and the SES/achievement slope within his/her classroom? Do the means and slopes of effective teachers who differ only in their orientation to delivering instruction to their students (i.e., equity vs. quality) vary?

Another issue which is inextricably tied to the issues of equity and excellence is the matter of how achievement is defined. Since all of the judgments made about these two issues are based on comparisons of one group to the other,

perhaps more thought should be given to what types of skills are being utilized to make these comparisons. For the most part, school and teacher effectiveness research (including the present study) have used the most convenient measures of student attainment - norm-referenced or state criterion-referenced examinations. These instruments, for the most part, are multiple-choice and product-oriented. It could be that the effects of the kinds of teacher behaviors that make the most difference in terms of educational attainment are not adequately captured with such an instrument. Although more costly to gather, it seems the use of more process oriented evaluation data is fundamental for deepening our understanding of what good teachers do. Once this type of data had been gathered, comparisons of results could be made. It is likely that the models for SES/product-oriented evaluation slopes and SES/process-oriented evaluation slopes would be very different.

Methodologically, the present study addresses some questions about the uses of HLM as a viable tool for classroom research and raises others. Although most of the results of this project echo those found with other methodologies, it could be asserted that because HLM more accurately models the nested structure of the data, its findings are more theoretically satisfying. However, it is difficult to envision a study seeking to measure similar teacher behaviors which would not encounter the same pattern

of intercorrelation. Investigations must be made which can offer some guidance as to the effects of significant correlation among Level-2 predictors and what strategies, if any, should be employed to deal with these effects in the event that they are a threat to the validity of results generated.

Methodological concerns notwithstanding, the results of this project are encouraging. Taken together with other studies which have noted that school effects explain 8-15% of the variance in individual student achievement (Bosker & Scheerens, 1989; Reynolds, 1992), there is clear evidence that given enough time in effective educational settings, a child's education attainment may be impacted significantly. Further, the unprecedented findings of the present study concerning the link between SES and achievement give evidence that this relationship is not one which is impervious to effective teaching. Most encouraging, however, is the fact that the variables investigated may all, in some sense, be manipulated. That is, it is within the power of those involved with learners on a day to day basis to improve and/or acquire the behaviors necessary for optimizing each child's opportunity to succeed academically.

REFERENCES

- Aitkin, M., Anderson, D., & Hinde, J. (1981). Statistical modelling of data on teaching styles (with discussion). Journal of the Royal Statistical Society, Series A, 144(4), 419-461.
- Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. Journal of the Royal Statistical Society, Series A, 149(1), 1-43.
- Anderson, L.M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. The Elementary School Journal, 79, 193-223.
- Armor, D. J. (1972). School and family effects on black and white achievement: A reexamination of the USOE data. In F. Mosteller & D. P. Moynihan (Eds.), On Equality of Educational Opportunity (pp. 168-229). New York: Vintage Books.
- Averch, H., Carroll, S.J., Donaldson, T., Kiesling, H.J., & Pincus, J. (1972). How effective is schooling? A critical review and synthesis of research findings (R-956-PCSF/RC). Santa Monica, CA: The RAND Corporation.
- Barr, R. & Dreeben, R. (1983). How schools work. Chicago: University of Chicago Press.
- Borg, W.R., & Gall, M.D. (1989). Educational research. (5th ed.). New York: Longman.
- Borich, G. (1986). Paradigms of teacher effectiveness research: Their relationship to the concept of effective teaching. Education and Urban Society, 18, 143-67.
- Bosker, R.J. & Scheerens, J. (1989). Issues in the interpretation of the results of school effectiveness research. International Journal of Educational Research, 13, 741-751.
- Boyd, W.L. (1987). Public education's last hurrah? Schizophrenia, amnesia and ignorance in school politics. Educational Evaluation & Policy Analysis, 9, 85-100.
- Brookover, W., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). Schools, social systems and student achievement: Schools can make a difference. New York: Praeger.

- Brophy, J. E. (1979). Teacher behavior and its effects. Journal of Educational Psychology, 71, 733-750.
- Brophy, J.E. (1982). Successful teaching strategies for the inner-city child. Phi Delta Kappan, 63, 527-530.
- Brophy, J. E. (1983). Classroom organization and management. Elementary School Journal, 83(4), 265-286.
- Brophy, J.E. & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), Handbook of research on teaching. New York: Macmillan.
- Brophy, J.E. & Evertson, C. (1976). Learning from teaching: A developmental perspective. Boston: Allyn and Bacon.
- Bryk, A.S., Holland, P., Lee, V., & Carriedo, R. (1984). Effective Catholic schools: An exploration. Washington, DC: National Catholic Educational Association.
- Bryk, A.S. & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101, 147-158,
- Bryk, A.S. & Raudenbush, S.W. (1989). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. In D. Bock (Ed.) Multilevel Analysis of Educational Data. San Diego, CA: Academic Press, Inc.
- Bryk, A.S. & Raudenbush, S.W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage Publications, Inc.
- Bryk, A.S., Raudenbush, S.W., Seltzer, M., & Congdon, R.T. (1986). An introduction to HLM: Computer program and user's guide. University of Chicago, Department of Education.
- Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. In D. C. Berliner (Ed.), Review of Research in Education, Vol. 8, (pp. 158-233). Washington, D.C.: American Educational Research Association.

- Burstein, L. (1982). Analysis of educational effects from a multilevel perspective: Disentangling between-and within-class relationships in mathematics performance. (Report No. CSE-R-172). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation. (ERIC Document Reproduction Service No. ED 228 311).
- Burstein, L., Kim, K.S. & Delandshere, G. (1989). Multilevel analysis of systematically varying slopes: Issues, alternative and consequences. In R.D. Bock (Ed.) Multilevel analysis of educational data. (pp.233-276). New York: Academic Press.
- Burstein, L., Miller, M.D. & Linn, R.L. (1981). The use of within-group slopes as indices of group outcomes. CSE Report No. 171, University of California, Los Angeles: Center for the Study of Evaluation.
- Coker, H., Medley, D., & Soar, R. (1980). How valid are expert opinions about effective teaching? Phi Delta Kappan, 62, 131-134, 149.
- Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). On equality of educational opportunity. Washington, DC: United States Department of Health, Education, and Welfare.
- Coleman, J.S., Hoffer, T., & Kilgore, S.B. (1982). High school achievement: Public, Catholic, and other private schools compared. New York: Basic.
- Corno L. (1979). Classroom instruction and the matter of time. In D.L. Duke's (Ed.), Classroom Management. Chicago: The University of Chicago Press, 245-280.
- Cronbach, L.J. (1976). Research on classrooms and schools: Formulation of questions, design, and analysis. Occasional paper: Stanford Evaluation Consortium.
- Crone, L (1992). The methodological issues of variance in teacher behavior and student achievement: The relationship of variance to school and teacher effectiveness. Unpublished Doctoral dissertation, Louisiana State University.
- DeLeeuw, J. & Kreft, I. (1986). Random coefficient models for multilevel analysis. Journal of Educational Statistics, 11, 57-85.

- Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM Algorithm (with discussion). The Royal Statistical Society, Series B, 39, 1-8.
- Doyle, W. (1984). How order is achieved in the classroom: An interim report. Journal of Curriculum Studies, 16(3), 259-277.
- Doyle, W. (1985). Recent research on classroom management: Implications for teacher preparation. Journal of Teacher Education, 3, 31-35.
- Edmonds, R.R. (1979a). Effective schools for the urban poor. Educational Leadership, 37, 15-24.
- Edmonds, R.R. (1979b). Some schools work and more can. Social Policy, 9(2), 28-32.
- Edmonds, R.R., and Frederiksen, J.R. (1978). Search for effective schools: The identification and analysis of city schools that are instructionally effective for poor children. Cambridge: Harvard University, Center for Urban Studies.
- Efron, B., & Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. Journal of the American Statistical Association, 74, 311-19.
- Emans, R.L. & Milburn, C.M. (1989). The knowledge base of teaching: A review and commentary of process-product education. Vermillion, SD: The University of South Dakota, School of Education. (ERIC Document Reproduction Service No. ED 309 137)
- Emmer, E., Evertson, C., & Anderson, L. (1980). Effective classroom management at the beginning of the school year. Elementary School Journal, 80, 219-231.
- Fitz-Gibbon, C.T. (1991). Multilevel modelling in an indicator system. In S.W. Raudenbush & J.D. Willms (Eds.), Schools, classrooms, and pupils: International studies of schooling from a multilevel perspective (pp. 67-83). Orlando, FL: Academic Press.
- Flanders, N. (1970). Analyzing teacher behavior. Reading, MA: Addison-Wesley.
- Gage, N.L. & Needels, M.C. (1989). Process-product research on teaching: A review of criticisms. The Elementary School Journal, 89, 253-300.

- Geske, T.G. & Teddlie, C. (1990). Research on school productivity: Where do we stand? In P. Reyes (Ed.), Employee Commitment, Performance, & Productivity in Educational Organizations. Newbury, CA: Sage.
- Goldstein, H.I. (1986). Efficient statistical modeling of longitudinal data. Annals of Human Biology, 13, 129-142.
- Good, T. L. (1983). Classroom research: A decade of progress. Educational Psychologist, 18(3), 127-144.
- Good, T.L. & Brophy, J. E. (1987). Looking in classrooms. (4th ed.) New York: Harper and Row.
- Good, T.L., & Grouws, D. (1975). Process-product relationship in fourth-grade mathematics classrooms. (Final Report: National Institute of Education Grant NIE-G--3-0123). Columbia: University of Missouri, College of Education.
- Good, T., Grouws, D. & Ebmeier, M. (1983). Active mathematics teaching. New York: Longman.
- Good, T., Ebmeier, M. & Beckerman, T. (1978). Teaching mathematics in high and low SES classrooms: An empirical comparison. Journal of Teacher Education, 29, 85-90.
- Greer, C. (1972). The great school legend. New York: Penguin Books.
- Hallinger, P., & Murphy, J. (1985, March-April). Instructional effectiveness and school socioeconomic status: Is what's good for the goose, good for the gander? Paper presented at the meeting of the American Educational Research Association, Chicago, IL.
- Hanushek, E.A. (1972). Education and race. Lexington, MA: D.C. Heath.
- Hanushek, E.A. (1986). The economics of schooling: Production and efficiency in public schools. Journal of Economic Literature, 24, 1141-1177.
- Hanushek, E.A., & Kain, J. F. (1972). On the value of equality of educational opportunity as a guide to public policy. In F. Mosteller & D. P. Moynihan (Eds.), On Equality of Educational Opportunity (pp. 116-145). New York: Vintage Books.

- Hedges, L.V. (1982). Fitting continuous models to effect size data. Journal of Educational Statistics, 7, 245-270.
- Karweit, N. (1983). Time on task: A research review (Report No. 332). Baltimore, MD: Johns Hopkins University, Center for Social Organization for Schools.
- Katzman, M. T. (1971). The political economy of urban schools. Cambridge, MA: Harvard University Press.
- Kennedy, E. (1990). A longitudinal study of the relationship of social class background to academic achievement for a cohort of public primary school students in South Carolina. Unpublished Doctoral dissertation, University of South Carolina.
- Kennedy, E., Teddlie, C., & Stringfield, S. (1991, January). A multilevel analysis of Phase II of The Louisiana School Effectiveness Study. Paper presented at the meeting of the International Congress for School Effectiveness and School Improvement, Cardiff, Wales UK.
- Kennedy, E., Stringfield, S. & Teddlie, C. (1993). Schools do make a difference. In C. Teddlie & S. Stringfield (Eds.), Schools do make a difference: Lessons learned from a 10-year study of school effects. (pp. 15-26). New York: Teachers College Press.
- Klitgaard, R.K., & Hall, S. R. (1974). Are there unusually effective schools? Journal of Human Resources, 7, 90-106.
- Kounin, J. S. (1970). Discipline and group management in classrooms. New York: Holt, Rinehart, and Winston.
- Lang, M.H. (1991). Effective school status: A methodological study of classification consistency. Unpublished Doctoral dissertation, Louisiana State University.
- Larivee, B., & Algina, J. (1983). Identification of teaching behaviors which predict success for mainstreamed students. Paper presented at the annual meeting of the American Educational Research Association, Montreal.
- Lee, V.E. (1986, April). Multilevel causal models for social class and achievement. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

- Levine, D.U. & Lezotte, L.W., (1990). Unusually effective schools: A review and analysis of research and practice. Madison, WI: University of Wisconsin, National Center for Effective Schools Research and Development.
- Levin, H. M. (1970). A new model of school effectiveness. In Do Teachers Make a Difference? (pp. 55-78). Washington DC: U.S. Department of Health, Education and Welfare.
- Lewis-Beck, M.S. (1987). Applied regression: An introduction. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-022. Beverly Hills and London: Sage Publications.
- Lindley, D.V., & Smith, A.F.M. (1972). Bayes estimates for the linear model. Journal of the Royal Statistical Society, Series B, 34, 1-41.
- Louisiana Department of Education (1989). Louisiana educational assessment program annual report. Baton Rouge, Louisiana Department of Education (1991).
- Mason, W.M., Hauser, R.M., Kerckhoff, A.C. & Poss, S.S. (1976). Models of response error in student reports of parental socioeconomic characteristics. In W.H. Sewell, R.M. Hauser & D. L. Fetherman (Eds.), Schooling and Achievement in American Society. New York: Academic Press.
- Mason, W.M., Wong, G.Y., & Entwisle, B. (1984). Contextual analysis through the multi-level linear model. In S. Leinhardt (Ed.), Sociological Methodology (pp. 72-103). San Francisco: Jossey-Bass.
- Medley, D. (1977). Teacher competence and teacher effectiveness: A review of process-product research. Washington, DC: American Association of Colleges for Teacher Education.
- Miller, S.K., & Yelton, B. (1987, April). Correlates of achievement in affluent effective schools. Paper presented at the meeting for the American Educational Research Association, Washington, D.C.
- Morris, C. (1983). Parametric empirical Bayes inference: Theory and applications. Journal of the American Statistical Association, 78, 47-65.

- Mortimore, P., Sammons, P., Stoll, L. Lewis, D. & Ecob, R. (1988a). School matters. Berkley: University of California Press.
- Mortimore, P., & Sammons, P. (1987). New evidence on effective elementary schools. Educational Leadership, 45(1), 4-8.
- Mosteller, F. & Moynihan, D.P. (Eds.) (1972). On equality of educational opportunity. New York: Vintage Books.
- Murnane, R.J. (1975). The impact of school resources on the learning of children. Cambridge, MA: Ballinger Publishing Co.
- Murnane, R. J. & Nelson, R. R. (1984). Production and innovation when techniques are tacit. Journal of Economic Behavior and Organization, 5, 353-373.
- Olson L. (1988, March 2). Teaching's knowledge base seen still elusive. Education Week, 7.
- Purkey, S.C. & Smith, M.W. (1983). Effective schools - A review. Elementary School Journal, 83, 427-52.
- Raudenbush, S.W. & Bryk, A.S. (1986). A hierarchical model for studying school effects. Sociology of Education, 59, 1-17.
- Raudenbush, S.W. & Bryk, A.S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. Review of Research in Education, 15, 423-475.
- Raudenbush, S.W. & Bryk, A.S. (1989). Quantitative models for estimating teacher and school effectiveness. In D. Bock (Ed.), Multilevel Analysis of Educational Data. San Diego, CA: Academic Press, Inc.
- Reynolds, D. (1992). School effectiveness and school improvement. In J. Bashi & Z. Sass (Eds.), School effectiveness and improvement: Selected proceedings of the Third International Congress for School Effectiveness (pp. 67-87). Jerusalem, Israel: Magnes Press.
- Rogoff, B. & Wertsch, J. (Eds.) (1984). Children's learning in the "zone" of proximal development. San Francisco, CA: Jossey-Bass.

- Rosenshine, B. V. (1976). Classroom instruction. In N.L. Gage (Ed.), The Psychology of Teaching Methods. Chicago: The University of Chicago Press, 335-71.
- Rosenshine, B. & Furst, N.F. (1973). The use of direct observation to study teaching. In R.N.W. Travers' (Ed.), Second Handbook of Research on Teaching. Chicago: Rand McNally, 122-83.
- Rouk, U. (1979). What makes an effective teacher? St. Louis, MO: Educational Research and Development Report of CEMREL, Inc.
- Rutter, M., Maughan, B., Mortimore, P., Ouston, J., and Smith, A. (1979). Fifteen thousand hours: Secondary schools and their effects on children. Cambridge, MA: Harvard University Press.
- SAS Institute, Inc. (1985). SAS User's Guide: Statistics. Version 5 Edition. Cary, NC.
- Smith, L. & Land, M. (1981). Low-inference verbal behaviors related to teacher clarity. Journal of Classroom Interaction, 17, 37-42.
- Soar, R. S. (1968). Optimum teacher-pupil interaction for pupil growth: Final report. Gainesville, FL: College of Education, University of Florida.
- Soar, R. S., & Soar, R. M. (1979). Emotional climate and management, In P. Peterson and H. Walberg (Eds.), Research on Teaching: Concepts, findings, and implications. Berkeley, CA: McCutchan.
- Soloman, D. & Kendall, A. J. (1976). Individual characteristics and children's performance in varied educational settings. Rockville, MD: Montgomery County Public Schools.
- St. John, N. (1970). The validity of children's reports of their parents' educational level: A methodological note. Sociology of Education, 43, 255-269.
- Stallings, J. (1980). Allocated learning time revisited, or beyond time on task. Educational Researcher, 9, 11-15.
- Stallings, J., Cory, R., Fairweather, J., & Needels, M. (1977). Early Childhood Education Classroom Evaluation, Menlo Park, CA: SRI International.

- Stallings, J. & Kaskowitz, D. (1974). Follow through classroom observation evaluation, 1972-1973. Office of Education Contract. OEC-0-852248-4633(100).
- Stallings, J., Needels, M. & Stayrook, N. (1979). How to change the process of teaching basic reading skills in secondary schools. Menlo Park, CA: SRI International.
- Stringfield, S., Teddlie, C., & Suarez, S. (1985). Classroom interaction in effective and ineffective schools: Preliminary results from Phase III of the Louisiana School Effectiveness Study. Journal of Classroom Interaction, 20, 31-37.
- Tate, R.L (1988). School comparisons with hierarchical linear models: An illustration. (Tech. Rep. No. 500151648) Tallahassee, FL: Florida State University.
- Teddlie, C., Falkowski, C., Stringfield, S., Desselle, S., and Garvue, R. (1984). The Louisiana school effectiveness study: Phase Two, 1982-1984. Baton Rouge, LA: Louisiana State Department of Education, 273 pages. Library of Congress call number: LA, 295, L877, 1984. (ERIC Document Reproduction Service No. ED 250 362).
- Teddlie, C., Kirby, P.C., & Stringfield, S. (1989). Effective versus ineffective schools: Observable differences in the classroom. American Journal of Education, 97, 221-236.
- Teddlie, & Stringfield (1993). Schools do make a difference: Lessons learned from a 10-year study of school effects. New York: Teachers College Press.
- Teddlie, C., Virgilio, I., & Oescher, J. (1990). Development and validation of the Virgilio Teacher Behavior Instrument. Educational and Psychological Measurement, 50 (2), 421-430.
- Tobin, K., & Capie, W. (1982). Relationships between classroom process variables and middle-school science achievement. Journal of Educational Psychology, 74, 441-454.
- Virgilio, I. (1987). An examination of the relationships among school effectiveness in elementary and junior high schools. Unpublished Doctoral dissertation, University of New Orleans.

- Virgilio, I., Teddlie, C., and Oescher, J. (1991). Variance and context differences in teaching at differentially effective schools. School Effectiveness and School Improvement, 2(2), 152-168.
- Weber, G. (1971). Inner-city children can be taught to read: Four successful schools. Washington, D.C.: Council for Basic Education.
- White, K.R. (1982). The relation between socioeconomic status and academic achievement. Psychological Bulletin, Vol. 91, 3, 461-481.
- Wimpelberg, R.K., Teddlie, C. & Stringfield, S. (1989). Sensitivity to context: The past and future of effective schools research. Educational Administration Quarterly, 25, 82-107.
- Wright, C. & Nuthall, G. (1970). Relationships between teacher behaviors and pupil achievement in three experimental elementary science lessons. American Educational Research Journal, 7, 477-491.

APPENDIX A

SCORING KEY AND SHEET

FOR

STALLINGS CLASSROOM SNAPSHOT

PLEASE NOTE

Copyrighted materials in this document have not been filmed at the request of the author. They are available for consultation, however, in the author's university library.

128-129

131-133

University Microfilms International

APPENDIX B
SCORING KEY AND SHEET
FOR
VIRGILIO TEACHER BEHAVIOR INVENTORY

APPENDIX C
STUDENT QUESTIONNAIRE

Name _____, _____
Last First

Your class has been chosen to be part of our research at LSU. These questions help us to get to know you a little better. Please answer the questions very carefully.

1. Please write your teacher's name.

2. Are you a girl or a boy? (circle the number of the correct answer)?

girl - 1
boy - 2

3. What is your race?

Black - 1
White - 2
Hispanic - 3
American Indian - 4
Asian - 5

4. How much education does your father have?

Finished middle school	- 1
Some high school	- 2
Finished high school	- 3
Some college	- 4
Finished college	- 5
Went to graduate school after college	- 6

5. How much education does your mother have?

Finished middle school	- 1
Some high school	- 2
Finished high school	- 3
Some college	- 4
Finished college	- 5
Went to graduate school after college	- 6

6. How many brothers and sisters do you have?

7. What is your mother's occupation? (Where does she work and what does she do on her job?)

8. What is your father's occupation? (Where does he work and what does he do on his job?)

VITA

Leslie Sanford Arceneaux received her bachelor's degree in Elementary Education from Louisiana State University in 1986 and went on to complete her Master's in Reading in 1988. During her doctoral work she has made presentations at several national conferences as well as having those papers published as part of meeting symposia. In the several years before the completion of her doctorate, Dr. Arceneaux has taught at the elementary level as a kindergarten, second grade, and resource teacher. She has also instructed at the university level - first as Assistant Professor of Education at McNeese State University in Louisiana, then as Assistant Professor of Education in the School of Business and Professional Studies at Lambuth University in Jackson, Tennessee.

At present, Dr. Arceneaux resides in Jackson, Tennessee with her husband, Shannon, and her son, Garrett Ross. Although taking a leave from teaching to be a full-time mother, she is still involved in the business of educating children through her work as a frequent lecturer at inservice training for local teachers.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Leslie Sanford Arceneaux

Major Field: Education

Title of Dissertation: The Influence of Teacher Behavior on the Distribution of Achievement in the Classroom: An Application of the Hierarchical Linear Model

Approved:

Engene Kennedy
Major Professor and Chairman

Daniel Poehl
Dean of the Graduate School

EXAMINING COMMITTEE:

Abbas Tashakk

Charles Tidd

Ferry R. Seshu

William L. Stefan

Brian H. Bernstein

Date of Examination:

October 28, 1993
